SMALL SPHERE DISTRIBUTIONS AND RELATED TOPICS IN DIRECTIONAL STATISTICS

by

ByungWon Kim

B.S. in Statistics, Korea University, Seoul, 2008

M.S. in Statistics, Korea University, Seoul, 2010

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

ByungWon Kim

It was defended on

June 20th 2018

and approved by

Dr. Sungkyu Jung, Department of Statistics

Dr. Yu Cheng, Department of Statistics

Dr. Zhao Ren, Department of Statistics

Dr. Stewart Anderson, Department of Biostatistics

Dissertation Director: Dr. Sungkyu Jung, Department of Statistics

ABSTRACT

SMALL SPHERE DISTRIBUTIONS AND RELATED TOPICS IN DIRECTIONAL STATISTICS

ByungWon Kim, PhD

University of Pittsburgh, 2018

This dissertation consists of two related topics in the statistical analysis of directional data. The research conducted for the dissertation is motivated by advancing the statistical shape analysis to understand the variation of shape changes in 3D objects.

The first part of the dissertation studies a parametric approach for multivariate directional data lying on a product of spheres. Two kinds of concentric unimodal-small subsphere distributions are introduced. The first kind coincides with a special case of the Fisher-Bingham distribution; the second is a novel adaption that independently models horizontal and vertical variations. In its multi-subsphere version, the second kind allows for correlation of horizontal variations over different subspheres. For both kinds, we provide new computationally feasible algorithms for simulation and estimation, and propose a large-sample test procedure for several sets of hypotheses. Working as models to fit the major modes of variation, the proposed distributions properly describe shape changes of skeletally-represented 3D objects due to rotation, twisting and bending. In particular, the multi-subsphere version of the second kind accounts for the underlying horizontal dependence appropriately.

The second part is a proposal of hypothesis test that is applicable to the analysis of principal nested spheres (PNS). In PNS, determining which subsphere to fit, among the geodesic (great) subsphere and non-geodesic (small) subsphere, is an important issue and it is preferred to fit a great subsphere when there is no major direction of variation in the directional data. The proposed test utilizes the measure of multivariate kurtosis. The change

of the multivariate kurtosis for rotationally symmetric distributions is investigated based on modality. The test statistic is developed by modifying the sample kurtosis. The asymptotic sampling distribution of the test statistic is also investigated. The proposed test is seen to work well in numerical studies with various data situations.

TABLE OF CONTENTS

1.0 INTRODUCTION 1

2.0 SMALL SPHERE DISTRIBUTIONS FOR DIRECTIONAL DATA . . 4

2.1 Introduction 4

2.2 Parametric small-sphere models 8

2.2.1 Two classical distributions on $\mathbb{S}^{p-1}$ 9

2.2.2 Small-sphere distributions of the first and second kind 10

2.2.3 Multivariate extensions 13

2.2.4 Random data generation 16

2.3 Maximum likelihood estimation 17

2.3.1 Estimation for S1 and iMS1 models 17

2.3.2 Estimation for S2, iMS2 and MS2 21

2.4 Testing hypotheses 23

2.5 Numerical studies 24

2.5.1 Estimation of small-circles 24

2.5.2 Estimation of horizontal dependence 27

2.5.3 Detecting overfitting in an isotropic case 28

2.6 Analysis of s-rep data 29

2.6.1 Modeling rotationally-deformed ellipsoids via s-reps 30

2.6.2 Data preparation 30

2.6.3 Inference on the bending axis 32

2.6.4 Inference on horizontal dependence 32

2.7 Human knee gait analysis 34

2.8 Technical Details and Supporting Materials 36
2.8.1 Proofs of lemmas and propositions 36
2.8.2 Convergence of algorithm for S1 estimation 38
2.8.3 Null distributions and empirical powers of tests 38
2.8.4 Additional simulation results 41
2.8.5 Associations among s-rep spokes 45
2.8.6 Goodness of fit 50
3.0 TEST OF MODALITY FOR ROTATIONALLY SYMMETRIC DISTRIBUTIONS BY MULTIVARIATE KURTOSIS 55
3.1 Introduction 55
3.2 Backgrounds 57
3.2.1 Sequential tests for preventing overfitting in PNS 57
3.2.2 Bayesian information criterion in PNS 59
3.2.3 Folded normal test 60
3.3 Test of rotational symmetry by tangent space likelihood ratio test 61
3.4 Test of modality for rotationally symmetric distributions 63
3.4.1 The measure of multivariate kurtosis 63
3.4.2 The multivariate kurtosis for rotationally symmetric distributions . . . 64
3.4.3 Test statistic and its asymptotic sampling distribution 71
3.4.3.1 Modified sample kurtosis for rotationally symmetric distributions 71
3.4.3.2 Proposed test procedure 72
3.4.3.3 Asymptotic sampling distribution of the modified sample kurtosis 73
3.5 Numerical studies 74
3.6 Appendix: The geodesic ball uniform on $\mathbb{S}^2$ 77
3.7 Appendix: Technical details 80
3.7.1 Proof of Lemma 5 80
3.7.2 Proof of Propositoin 6 82
3.7.3 Details in proof of Proposition 8 84
3.7.3.1 Approximation of the modified sample kurtosis in (3.14) . . . 84
3.7.3.2 Calculation of the expected values and variances of R_i and T_i 85

BIBLIOGRAPHY 87

LIST OF TABLES

1 Key features of the newly proposed models 8
2 Small-circle estimation performances for univariate S2 data 26
3 Small-circles estimation performances for bivariate S2 data 27
4 Concentration and association parameter estimation performances 28
5 Small-circle estimation performances for univariate signal-plus-noise data . . 47
6 Small-circles estimation performances for bivariate signal-plus-noise data . . . 47
7 Table of performance ($n = 50$) 78
8 Table of performance ($n = 200$) 78

LIST OF FIGURES

1 S-reps of hippocampi in the human brain . . . 3
2 Spoke directions on the unit sphere . . . 3
3 Toy examples: Small-circle-concentrated data . . . 5
4 The S1 densities on $\mathbb{S}^2$ . . . 12
5 Random samples from proposed models . . . 18
6 Simulation data examples from univariate S2 . . . 25
7 Examples showing degrees of the small-circle feature . . . 29
8 Analysis of s-reps of bent ellipsoid by MS2 . . . 31
9 Analysis of human knee gait data by MS2 . . . 35
10 Convergence of S1 estimation algorithm . . . 38
11 Q-Q envelope plots for distributions of test statistics . . . 40
12 Empirical powers of the test of association . . . 42
13 Data examples in test of great-sphere . . . 43
14 Data examples in test for BM distribution . . . 44
15 Simulation data examples from bivariate MS2 . . . 46
16 Power of test of association for bi- and tri-variate directions . . . 49
17 Data examples for which MS2 test of association is superior ($K = 2$) . . . 51
18 Data examples for which MS2 test of association is superior ($K = 3$) . . . 52
19 Goodness-of-fit analysis for fitting s-rep data . . . 54
20 Random data examples of interest . . . 56
21 Description of the ring uniform distribution . . . 66
22 The population multivariate kurtosis for the ball and ring uniform . . . 68

23 The population measure of multivariate kurtosis for a convolution and mixture 70
24 The change of population multivariate kurtosis 71
25 Random data examples for numerical studies 76

1.0 INTRODUCTION

Directional statistics is to analyze angular or directional observations which are represented by unit vectors in 2D or 3D space. Since the natural sample space of directional vectors is not Euclidean space, the standard statistical methods developed using Euclidean geometry for analyzing vector-valued observations can not be directly applied to directional data. To take into account the structure of the (non-Euclidean) sample space, a substantial amount of literature is available for directional data. See Mardia and Jupp (2000) for a good introduction.

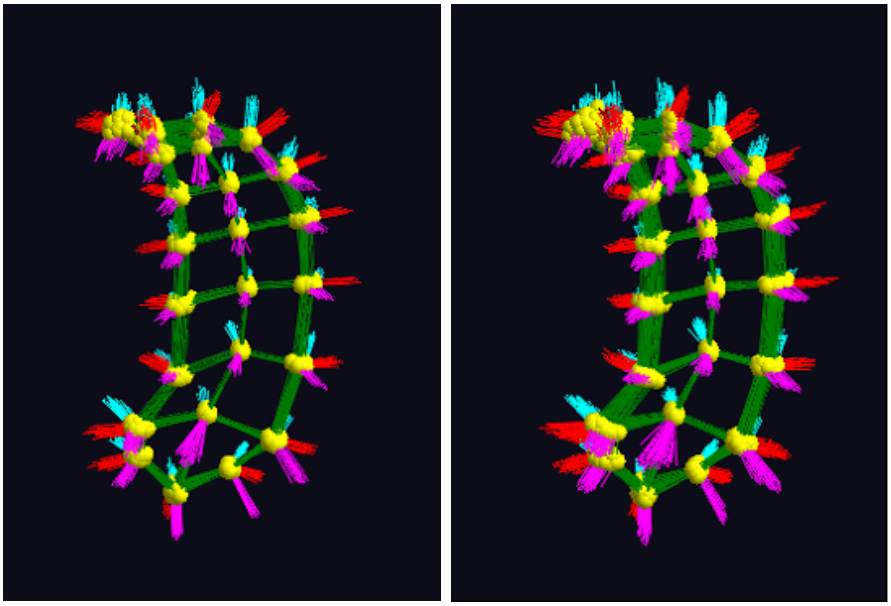
This thesis aims to contribute to the field of directional statistics by introducing and investigating statistical models for directional observations which are distributed on the unit sphere. This work contains two separate parts. The first one is about parametric models and related estimation algorithms which are newly introduced in our work. The second is on the hypothesis testing procedure which is applied to the analysis of principal nested spheres (PNS) introduced in Jung et al. (2012).

In both parts, the fundamental ideas came from a data-analytic situation in providing statistical understanding of 'shape changes' of 3D objects, but the proposed methods are widely applicable to problems in shape analysis, image analysis, and directional data analysis. Figure 1 shows a data example motivating this work. The figure shows the skeletal representations (s-reps; Siddiqi and Pizer, 2008), modeling the hippocampus in human brain, obtained from 274 subjects. The subjects in this study are suffered from the first episode schizophrenia. As the figure illustrates, shape and size of the object are represented by spoke vectors (red, blue, purple lines) connecting the skeletal positions (yellow dots) to the boundary of the object. It is common in the analysis of s-reps that the spoke vectors are decomposed into direction vectors and their lengths, to distinguish the shape from the

size. More detailed description and available methods for modeling 3D objects are given in Section 2.1 and 2.6.

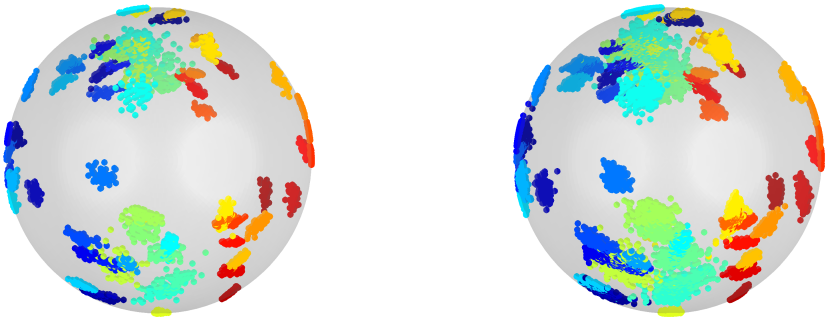
Figure 2 shows the directional vectors on the unit sphere from the hippocampi data. Some of these directional vectors are distributed along a non-straight curve. Capturing the major mode of variation of these directional vectors has been of interest for some time. In Chaper 2, we propose two new probability distributions for modeling directional data that are concentrated on a small circle, representing the non-geodesic curve. The proposed probability distributions are not in general rotationally symmetric, but have a unique mode on the non-geodesic curve. Since several directional vectors, that are associated with each other, are typically observed in s-reps data and in other applications (cf. Figure 2), we develop proper multivariate extensions to model dependencies among directions. Moreover, we propose efficient estimation algorithms for maximum likelihood estimations and introduce related large-sample tests with several important hypotheses.

In Chapter 3, a new hypothesis testing procedure is proposed, which is used to determine modality for rotationally symmetric distributions. The proposed test is specifically designed to improve the standard methods in the analysis of s-reps such as the composite principal nested spheres (CPNS; Pizer et al., 2013) and the polyspheres principal component analysis (PPCA; Eltzner et al., 2015). These methods are used to facilitate predictive analysis of s-reps by reducing the dimension of the abstract space of s-reps data, and also to represent the data in a lower-dimensional vector space. The abstract space of s-reps with K spokes is $\mathbb{R}^{K+1} \times \mathbb{S}^{3K-4} \times (\mathbb{S}^2)^K$, i.e., the Cartesian product of a vector space, a high-dimensional unit-sphere and K unit-spheres. For the hippocampi data given in Figure 1, the number of spokes is $K = 66$. To reduce the dimensions of a hypersphere $\mathbb{S}^{3K-4}$ and a polysphere (a product of spheres) $(\mathbb{S}^2)^K$ by capturing essential modes of variation, the PNS analysis is used in both CPNS and PPCA. The proposed testing procedure is used as an ad-hoc test in the procedure of the PNS to determine the radius of subspheres (great/small subspheres) to be fitted. Preliminary simulation study shows the proposed test has the highest power among all comparisons while keeping its size at the given level of significance.



(a) Control group (b) Treatment group

Figure 1: S-reps of hippocampi in the human brain



(a) Control group (b) Treatment group

Figure 2: Spoke directions on the unit sphere. Different colors indicate different spoke vectors.

2.0 SMALL SPHERE DISTRIBUTIONS FOR DIRECTIONAL DATA

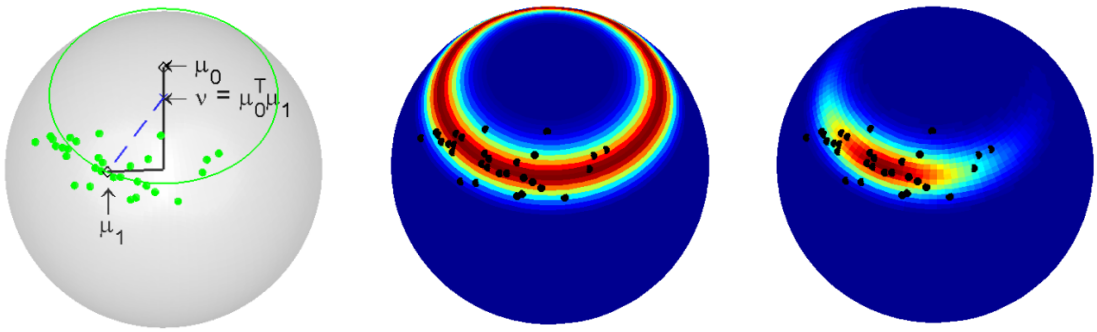
2.1 INTRODUCTION

In medical imaging, accurately assessing and correctly diagnosing shape changes of internal organs is a major objective of a substantial challenge. Shape deformations can occur through long-term growth or necrosis as well as by short-term natural deformations. In view of surgery and radiation therapy, it is important to model all possible variations of object deformations by both long- and short-term changes, in order to control the object's exact status and shape at treatment time. *Rotational deformations* such as rotation, bending, and twisting form a key sub-category of possible shape changes. For instance, shape changes of hippocampi in the human brain have been shown to mainly occur in the way of bending and twisting (Joshi et al., 2002; Pizer et al., 2013).

For the task of modeling 3D objects an abundance of approaches have been introduced. Closely related to our work are landmark-based shape models (Cootes et al., 1992; Dryden and Mardia, 1998; Kurtek et al., 2011) where a solid object is modeled by the positions of surface points, chosen either anatomically, mathematically or randomly. A richer family of models is obtained by attaching directions normal to the sampled surface points. More generally, in skeletal representations (Siddiqi and Pizer, 2008), an object is modeled by the combination of skeletal positions (lying on a medial sheet inside of the object) and spoke vectors (connecting the skeletal positions with the boundary of the object). In these models, describing the variation of rotational deformations can be transformed into a problem of exploring the motion of *directional vectors* on the unit two-sphere. As argued in Schulz et al. (2015), directional vectors representing rotational deformations tend to be concentrated on

small circles on the unit sphere; a toy data example in Fig. 3 shows a typical pattern of such observations.

For such s-rep data, spread out non-uniformly over several concentric small circles, to the best knowledge of the authors, there are neither parametric models nor inferential methods available. In order to fill this gap, in particular, to model *horizontal* (detailed below) dependence across different small subspheres, which are typical for s-rep data, we propose two new families and provide methods for estimation, simulation and statistical tests. To date, only for estimation of small circles, not involving horizontal dependence, though, there is only the nonparametric *least squares* (LS) method by Schulz et al. (2015) available. For the more simple task of estimating a single small circle, along which data is spread uniformly, there is the parametric family of *Bingham-Mardia* (BM) distributions by Bingham and Mardia (1978) available. We remove the uniformity constraint by adding a von Mises-Fisher term, giving either a special case of a Fisher Bingham distribution (Kent, 1982) or more subtly, a new family of distributions. For the former, while simulation and MLE methods are available (*cf.*, Hoff, 2009), for computational feasibility we adapt a saddlepoint approximation



(a) Data near a small circle (b) Bingham-Mardia (c) Proposed density

Figure 3: (a) Toy example showing observations (solid green) distributed near a small circle $\mathcal{C}(\mu, \nu)$. The heat maps of fitted Bingham-Mardia density (b) and the proposed small-sphere density of the first kind (c) are overlaid. Red: high density, blue: low density.

of Kume and Wood (2005). For the latter new one, in the case of small circles, we develop even faster numerical methods for simulation and estimation. For application to s-rep data we propose several multivariate extensions, in particular, in order to model horizontal dependence across different small subspheres. We show the usefulness of our new methods by analyzing s-rep data and comparing to the limited capabilities of LS and BM using methods derived ad-hoc. As mentioned, while LS and BM cannot model horizontal dependence, for comparison, we also derive a crude ad-hoc method to implement in *composite principal nested spheres* (CPNS) from Pizer et al. (2013) a test for horizontal dependence.

Let us now provide more detail. Throughout this paper, $\mathbb{S}^{p-1} = \{x \in \mathbb{R}^p \,|\, \|x\| = 1\}$ is the unit sphere in arbitrary dimension $p \geq 3$ and $\|x\| = (x^\top x)^{1/2}$ is the usual L_2-norm of the vector x. To precisely describe the targeted data situation, we define a $(p-2)$-dimensional *subsphere* of $\mathbb{S}^{p-1}$ as the set of all points equidistant from $\mu \in \mathbb{S}^{p-1}$, denoted by

$$\mathcal{C}(\mu, \nu) = \{x \in \mathbb{S}^{p-1} \,|\, \delta(\mu, x) = \arccos(\nu)\}, \quad \nu \in (-1, 1).$$

Here, $\delta(u, v) = \arccos(u^\top v)$ is the geodesic distance between $u, v \in \mathbb{S}^{p-1}$. The subsphere is called a *great subsphere* if $\nu = 0$ and a proper *small subsphere* if $\nu \neq 0$. Note that $\mathcal{C}(\mu, \nu) \subset \mathbb{S}^{p-1}$ is well-defined for all $p > 1$. For the special case of $p = 3$, $\mathcal{C}(\mu, \nu)$ is a circle, a *great circle* if $\nu = 0$, and a proper *small circle* if $\nu \neq 0$. To model the data in Fig. 3, one may naively use the BM distribution, which is a family of densities on $\mathbb{S}^2$ with a modal ridge along a small circle. However, typical observations we encountered in applications do not uniformly spread over the full circle, and the BM distribution does not fit well, as shown in Fig. 3(b). Moreover, when by a single observation multiple directional vectors are provided, that is, data are on a polysphere $(\mathbb{S}^2)^K$, to the knowledge of the authors, there is no tool available to date, to model dependencies between directions.

In this chapter, we propose two types of new distributional families for random directional vectors on $\mathbb{S}^{p-1}$, which we call *small-sphere distributions* of the first (S1) and second (S2) kind. If $p = 3$, the proposed distributions may be called *small-circle distributions*. These two distributional families are designed to have higher densities on $\mathcal{C}(\mu, \nu)$ and to have a unique mode on $\mathcal{C}(\mu, \nu)$. An example of a small-sphere density, fitted to the toy data is shown in Fig. 3(c). The new densities are natural extensions of the BM distribution with an

additional term explaining a decay from a mode. If the additional term is a von Mises-Fisher (vMF) density on $\mathbb{S}^{p-1}$, we obtain the S1, which is a subfamily of the general Fisher-Bingham distribution (Mardia, 1975; Mardia and Jupp, 2000). On the other hand, if the additional term is a vMF density on the subsphere ($\cong \mathbb{S}^{p-2}$), we obtain the S2 distribution, in which case the *horizontal* (inside the small subsphere) and *vertical* (orthogonal to the small subsphere) components of the directional vectors are independent of each other.

Several multivariate extensions of the new distributions to $(\mathbb{S}^{p-1})^K$, $K \geq 2$, are discussed as well. In particular, we show that a special case, called MS2, of our multivariate extensions is capable of modeling *dependent* random vectors. It has a straightforward interpretation, and we provide fast estimation of its parameters. This MS2 distribution is specifically designed with s-rep applications in mind. In particular, s-rep data from rotationally-deformed objects have directional vectors that are "rotated together," share a common axis of rotation, and are "horizontally dependent" (when the axis is considered to be vertically positioned). The component-wise independence of the S2 distributions plays a key role in this simple and interpretable extension. We discuss here likelihood-based parameter estimation and testing procedures of the multivariate distributions.

While the new distributions summarized in Table 1 contribute to the literature of directional distributions (Mardia and Jupp, 2000), the proposed estimation procedures for the S1, S2 and MS2 parameters can be thought of as a method of fitting small-subspheres to data, which has been of separate interest. Nonparametric least-squares type solutions for such problem dates back to Mardia and Gadsden (1977), Gray et al. (1980), and Rivest (1999). Jung et al. (2012) proposed recursively fitting small-subspheres in dimension reduction of directional and shape data. Pizer et al. (2013) proposed to combine separate small-circle fitting results in the analysis of s-rep data. In a similar spirit, Jung et al. (2011) and Schulz et al. (2015) also considered fitting small-circles in applications to s-rep analysis. In a simulation study, we show that our estimators provide smaller mean angular errors in small-circle fits than recent developments listed above.

The rest of this chapter is organized as follows. In Section 2.2, we introduce the proposed densities of the S1 and S2 distributions and discuss their multivariate extensions including the MS2 distribution. Procedures of obtaining random variates from the proposed distributions

	Relation among horizontal and vertical components		
	Dependent	Independent	
Univariate	S1	S2	
Multivariate (indep.)	iMS1	iMS2	
Multivariate (dep.)	×	GMS2	MS2 ($p-1=2$)
Simulation	Gibbs sampling	×	Exact sampling
Estimation	Approximate MLE	×	Approximate MLE
Hypothesis testing	Likelihood ratio	×	Likelihood ratio

Table 1: Key features of the newly proposed small-sphere distributions (top three rows) and methods (bottom three rows) developed in this paper. Items marked "×" are beyond the scope of this paper.

are also discussed. In Section 2.3, algorithms to obtain maximum likelihood estimators of the parameters are proposed and discussed. In Section 2.4, we introduce several hypotheses of interest and procedures of likelihood-ratio tests. Simulation studies demonstrating the performance of small-circle fitting, estimating dependency, and the power of the proposed test are contained in Section 2.5. In Sections 2.6 and 2.7 we demonstrate applications of the new multivariate distributions to analyze models that represent human organs and knee motions. In Section 2.8, we provide supporting details including proofs and additional numerical results.

2.2 PARAMETRIC SMALL-SPHERE MODELS

First we introduce two classical spherical densities, then we suitably combine them for our purposes.

2.2.1 Two classical distributions on $\mathbb{S}^{p-1}$

The von Mises-Fisher (vMF) distribution (Mardia and Jupp, 2000, p.168) is a fundamental unimodal and isotropic distribution for directions with density

$$f_{\text{vMF}}(x;\mu,\kappa) = \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{\Gamma(p/2)\mathcal{I}_{p/2-1}(\kappa)} \exp\{\kappa\mu^\top x\}, \quad x \in \mathbb{S}^{p-1}. \tag{2.1}$$

Here, Γ is the gamma function and $\mathcal{I}_v$ is the modified Bessel function of the first kind and order v. The parameter $\mu \in \mathbb{S}^{p-1}$ locates the unique mode with $\kappa \geq 0$ representing the degree of concentration.

The Bingham-Mardia (BM) distribution was introduced by Bingham and Mardia (1978) to fit data in $\mathbb{S}^2$ that cluster near a small circle $\mathcal{C}(\mu,\nu)$. For an arbitrary dimension $p \geq 3$, the BM density is given by

$$f_{\text{BM}}(x;\mu,\kappa,\nu) = \frac{1}{\alpha(\kappa,\nu)} \exp\{-\kappa(\mu^\top x - \nu)^2\}, \quad x \in \mathbb{S}^{p-1}, \tag{2.2}$$

where $\alpha(\kappa,\nu) > 0$ is the normalizing constant.

For our purpose of generalizing these distributions, we represent the variable $x \in \mathbb{S}^{p-1}$, $p \geq 3$, by spherical angles $\phi_1, \ldots, \phi_{p-1}$ satisfying $\cos\phi_1 = \mu^\top x$. Setting $s := \cos\phi_1 \in [-1,1]$ and $\phi := (\phi_2, \ldots, \phi_{p-1}) \in [0,\pi]^{p-3} \times [0,2\pi)$, the random vector (s,ϕ) following the von Mises-Fisher (2.1) or Bingham-Mardia (2.2) distribution has the respective density:

$$g_{\text{vMF}}(s,\phi;\kappa) = \left(\frac{\kappa}{2}\right)^{p/2-1} \frac{1}{\Gamma(p/2)\mathcal{I}_{p/2-1}(\kappa)} \exp\{\kappa s\}, \tag{2.3}$$

$$g_{\text{BM}}(s,\phi;\kappa,\nu) = \frac{1}{\alpha(\kappa,\nu)} \exp\{-\kappa(s-\nu)^2\}. \tag{2.4}$$

In consequence, for both distributions, s and ϕ are independent, and the marginal distribution of ϕ, which parametrizes a co-dimension 1 unit sphere $\mathbb{S}^{p-2}$, is uniform. In (2.3), the marginal distribution of s is a shifted exponential distribution truncated to $s \in [-1,1]$, while in (2.4) the marginal distribution of s is a normal distribution truncated to $s \in [-1,1]$. Both densities are isotropic, i.e. rotationally symmetric with respect to μ. The vMF density is maximal at the mode μ and decreases as the latitude ϕ_1 increases, while the BM density is uniformly maximal on the small-sphere $\mathcal{C}(\mu,\nu)$ and decreases as ϕ_1 deviates from $\arccos(\nu)$.

2.2.2 Small-sphere distributions of the first and second kind

The proposed small-sphere densities of the first and second kind on $\mathbb{S}^{p-1}$, for $x = (x_1, \ldots, x_p) \in \mathbb{S}^{p-1}$ with parameters $\mu_0, \mu_1 \in \mathbb{S}^{p-1}$, $\nu = \mu_0^\top \mu_1 \in (-1, 1)$, $\kappa_0 > 0$, $\kappa_1 > 0$, are given by

$$f_{\mathrm{S1}}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \frac{1}{a(\kappa_0, \kappa_1, \nu)} \exp\{-\kappa_0(\mu_0^\top x - \nu)^2 + \kappa_1 \mu_1^\top x\}, \tag{2.5}$$

$$f_{\mathrm{S2}}(x; \mu_0, \mu_1, \kappa_0, \kappa_1) = \frac{1}{b(\kappa_0, \kappa_1, \nu)} \exp\left\{-\kappa_0(\mu_0^\top x - \nu)^2 + \kappa_1 \frac{\mu_1^\top P_{\mu_0} x}{\sqrt{\mu_1^\top P_{\mu_0} \mu_1 x^\top P_{\mu_0} x}}\right\} \tag{2.6}$$

respectively, where $a(\kappa_0, \kappa_1, \nu)$ and $b(\kappa_0, \kappa_1, \nu)$ are normalizing constants. Here, P_{μ_0} denotes the matrix of orthogonal projection to the orthogonal *complement* of μ_0; $P_{\mu_0} = I_p - \mu_0 \mu_0^\top$, where I_p is the identity matrix. (In (2.6), we use the convention $0/0 = 0$.)

These distributions are well-suited to model observations that are concentrated near the small sphere $\mathcal{C}(\mu_0, \nu)$ but are not rotationally symmetric. The first kind (2.5) is a natural combination of the vMF (2.1) and BM (2.2) distributions. The parameter μ_0 represents the axis of the small sphere $\mathcal{C}(\mu_0, \nu)$, while μ_1 gives the mode of the distribution, which, by the definition of ν, is on the small sphere $\mathcal{C}(\mu_0, \nu)$. These parameters, μ_0, μ_1, ν, are illustrated in Fig. 3(a) for the $p = 3$ case. The parameter κ_0 controls the *vertical concentration* towards the small sphere (with an understanding that μ_0 is arranged vertically). In (2.5), κ_1 controls the isotropic part of the concentration around the mode, forcing the density to decay from μ_1.

The rationale for the second kind (2.6) is better understood with a change of variables. Let us assume for now that $\mu_0 = (1, 0, \ldots, 0)^\top$. For any $x = (x_1, \ldots, x_p)^\top \in \mathbb{S}^{p-1}$, write $s := x_1 = \mu_0^\top x$. If the spherical coordinate system $(\phi_1, \ldots, \phi_{p-1})$ as defined for (2.4) is used, then $s = \cos\phi_1$. The "orthogonal complement" of s is denoted by

$$y := (x_2, \ldots, x_p)/\sqrt{1 - s^2} \in \mathbb{S}^{p-2}, \tag{2.7}$$

where the vector y is obtained from the relation $P_{\mu_0} x / \|P_{\mu_0} x\| = (0, y) \in \mathbb{S}^{p-1}$. Similarly, define $\widetilde{\mu_1} \in \mathbb{S}^{p-2}$ as the last $p - 1$ coordinates of $P_{\mu_0}\mu_1 / \|P_{\mu_0}\mu_1\|$. Then the random vector

$(s, y) \in [-1, 1] \times \mathbb{S}^{p-2}$ from the S1 or S2 has densities

$$g_{\mathrm{S1}}(s, y; \mu_1, \kappa_0, \kappa_1) = \frac{1}{a(\kappa_0, \kappa_1, \nu)} \exp\left\{-\kappa_0(s-\nu)^2 + \kappa_1 \mu_1^\top \left(s, \sqrt{1-s^2}y\right)\right\}, \quad (2.8)$$

$$g_{\mathrm{S2}}(s, y; \mu_1, \kappa_0, \kappa_1) = \frac{1}{b(\kappa_0, \kappa_1, \nu)} \exp\left\{-\kappa_0(s-\nu)^2 + \kappa_1 \widetilde{\mu_1}^\top y\right\}, \quad (2.9)$$

respectively, for $s \in [-1, 1]$, $y \in \mathbb{S}^{p-2}$. The subtle difference is that for the first kind (2.8), the "vMF part" (the second term in the exponent) is not statistically independent from the "BM part", while it is true for the second kind (2.9). That is, s and y are independent only in the second kind. Accordingly, in (2.9), κ_1 controls the *horizontal concentration* towards the mode μ_1. The parameters μ_0, μ_1 and κ_0 of the second kind have the same interpretations as those of the first kind.

We use the notation $X \sim \mathrm{S1}(\mu_0, \mu_1, \kappa_0, \kappa_1)$ and $Y \sim \mathrm{S2}(\mu_0, \mu_1, \kappa_0, \kappa_1)$ for random directions $X, Y \in \mathbb{S}^{p-1}$ following small-sphere distributions of the first and second kind with parameters $(\mu_0, \mu_1, \kappa_0, \kappa_1)$, respectively. The proposed distributions are quite flexible and can fit a wide range of data. In Figure 4, we illustrate the S1 densities (2.5) with various values of the concentration parameters κ_0, κ_1. In all cases, the density is relatively high near the small circle $\mathcal{C}(\mu_0, \nu)$ and has the mode at $\mu_1 \in \mathcal{C}(\mu_0, \nu)$. Despite the difference in their formulations, the S2 densities (2.6) look similar to S1 densities for each fixed parameter-set.

Both distributions are invariant to rotation in the nullspace of (μ_0, μ_1):

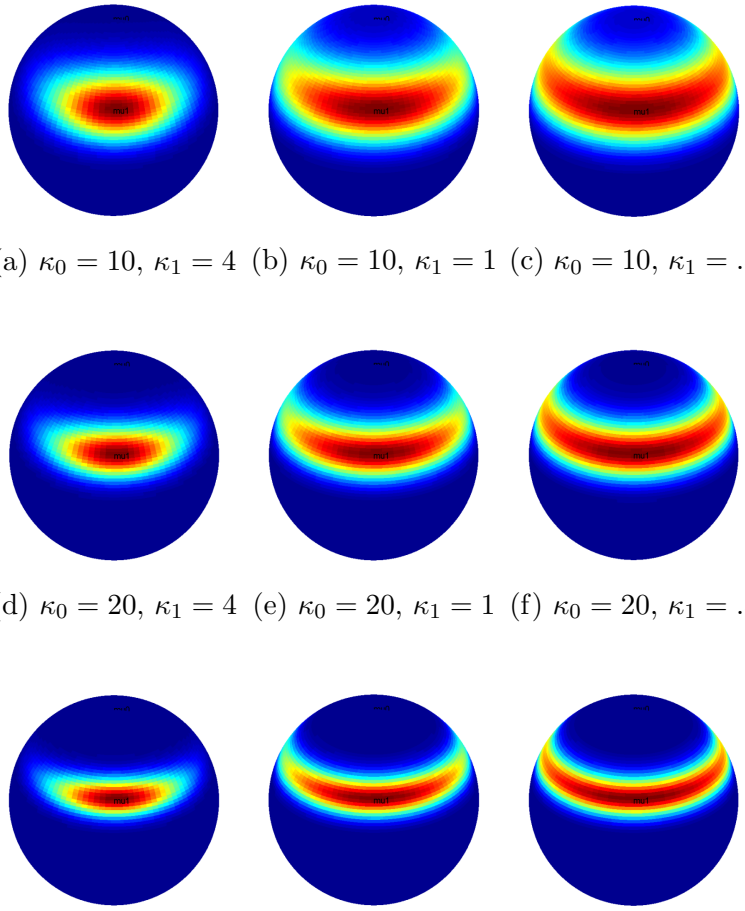
Proposition 1. *Let $X, Y \in \mathbb{S}^{p-1}$ be random directions with $X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$ and $Y \sim S2(\mu_0, \mu_1, \kappa_0, \kappa_1)$ and let B be a $p \times p$ orthogonal matrix.*

(i) X and BX (or Y and BY) have the same distribution if and only if $B\mu_0 = \mu_0$ and $B\mu_1 = \mu_1$.

(ii) $X \sim S1(-\mu_0, \mu_1, \kappa_0, \kappa_1)$ and $Y \sim S2(-\mu_0, \mu_1, \kappa_0, \kappa_1)$.

An example for the matrix B in Proposition 1(i) is the reflection matrix $B = I_p - 2UU^\top$, where $U = [u_3, \ldots, u_p]$ is such that $[u_1, \ldots, u_p]$ is a $p \times p$ orthogonal matrix whose column vectors u_1 and u_2 generate μ_0 and μ_1.

Remark 1. The S1 distribution is a special case of the Fisher-Bingham distribution (Mardia, 1975). Following the notation of Kent (1982), the S1 distribution may be labeled as a FB_6 *distribution*, in the special case of $p = 3$, emphasizing the 6-dimensional parameter space. In



(a) $\kappa_0 = 10$, $\kappa_1 = 4$ (b) $\kappa_0 = 10$, $\kappa_1 = 1$ (c) $\kappa_0 = 10$, $\kappa_1 = .5$

(d) $\kappa_0 = 20$, $\kappa_1 = 4$ (e) $\kappa_0 = 20$, $\kappa_1 = 1$ (f) $\kappa_0 = 20$, $\kappa_1 = .5$

(g) $\kappa_0 = 40$, $\kappa_1 = 4$ (h) $\kappa_0 = 40$, $\kappa_1 = 1$ (i) $\kappa_0 = 40$, $\kappa_1 = .5$

Figure 4: The S1 densities on $\mathbb{S}^2$ modeling non-isotropic small-circle distributions. High density (red), low density (blue). In all figures, μ_0 points to the north pole and μ_1 satisfies $\mu_0^\top \mu_1 = 1/2$. Rows and columns correspond to different choices of concentration parameters (κ_0, κ_1).

terms of the general parameterization of the Fisher-Bingham density (*cf.* Mardia and Jupp, 2000, p.174), we write $\gamma = 2\kappa_0\nu\mu_0 + \kappa_1\mu_1$ and $A = \kappa_0\mu_0\mu_0^\top$, so that the S1 density (2.5) is expressed as

$$f_{\mathrm{S1}}(x; \gamma, A) = \frac{1}{\alpha(\gamma, A)} \exp\{\gamma^\top x - x^\top A x\}, \tag{2.10}$$

where $\alpha(\gamma, A) = a(\kappa_0, \kappa_1, \nu) \exp\{\kappa_0 \nu^2\}$. This relation to the general Fisher-Bingham distribution facilitates random data generation and maximum likelihood estimation, shown later in Sections 2.2.4 and 2.3.1.

2.2.3 Multivariate extensions

The univariate small-sphere distributions (2.5–2.6) are now extended to model a tuple of *associated* random directions, $\mathbf{X} = (X_1, \ldots, X_K) \in (\mathbb{S}^{p-1})^K$. We confine ourselves to a special case where the marginal distributions of X_k have a common "axis" parameter μ_0, but relaxing this condition is straightforward. We begin by introducing multivariate small-sphere distributions for independent random directions, denoted by iMS1 and iMS2.

Independent extensions. Suppose that, in the K-tuple of random directions $\mathbf{X}$, each $X_k \in \mathbb{S}^{p-1}$ is marginally distributed as $\mathrm{S1}(\mu_0, \mu_k, \kappa_{0k}, \kappa_k)$. Throughout, we assume that $\nu_k = \mu_0^\top \mu_k \in (-1, 1)$ so that the underlying small spheres do not degenerate. If the components of $\mathbf{X}$ are mutually independent, then the joint density evaluated at $\mathbf{x} \in (\mathbb{S}^{p-1})^K$ is

$$f_{\mathrm{iMS1}}(\mathbf{x}) \propto \exp\left\{\boldsymbol{\Gamma}^\top \mathbf{x} - \mathbf{x}^\top \mathbf{A} \mathbf{x}\right\}. \tag{2.11}$$

Here, $\boldsymbol{\Gamma} = [\gamma_1, \ldots, \gamma_K]^\top$, where $\gamma_k = 2\kappa_{0k}\nu_k\mu_0 + \kappa_k\mu_k$, and $\mathbf{A} = \mathbf{K}_0 \otimes (\mu_0\mu_0^\top)$, where $\mathbf{K}_0 = \mathrm{diag}(\kappa_{01}, \ldots, \kappa_{0K})$. Each marginal density is of the form (2.10).

If each component is marginally distributed as $\mathrm{S2}(\mu_0, \mu_k, \kappa_{0k}, \kappa_k)$, then writing the density in terms of (s, y) as done for (2.9) facilitates our discussion. First, we decompose each x_k into $s_k = \mu_0^\top x_k \in [-1, 1]$ and $y_k \in \mathbb{S}^{p-2}$ as defined in (2.7). Further, we denote by $\widetilde{\mu_k}$ the scaled projection of μ_k as done for the univariate case. Then an independent multivariate extension for the S2 model can be expressed as the joint density of $\mathbf{s} = (s_1, \ldots, s_K)$ and $\mathbf{y} = (y_1, \ldots, y_K)$,

$$g_{\mathrm{iMS2}}(\mathbf{s}, \mathbf{y}) \propto \exp\left\{H^\top \mathbf{s} - \mathbf{s}^\top \mathbf{K}_0 \mathbf{s} + \mathbf{M}^\top \mathrm{vec}(\mathbf{y})\right\}, \tag{2.12}$$

where $H = (2\kappa_{01}\nu_1, \ldots, 2\kappa_{0K}\nu_K)$ and $\mathbf{M} = \mathrm{vec}(\kappa_1\widetilde{\mu_1}, \ldots, \kappa_K\widetilde{\mu_K})$ while $\mathrm{vec}(\cdot)$ denotes the column-wise vectorization of a matrix.

Vertical and horizontal dependence. Based on (2.11) and (2.12), we now contemplate dependent models. Obviously, if we allow in (2.11) nonzero offdiagonal entries of $\mathbf{A}$, then we obtain a dependent modification of the S1 model. With our applications in mind, however, we aim at modeling a specific structure of dependence that is natural to the variables $(\mathbf{s}, \mathbf{y})$ in (2.12).

If $s_1, \ldots, s_K$ are dependent, we speak of *vertical dependence*; if $y_1, \ldots, y_K$ are dependent, we speak of *horizontal dependence*. In practice, when we deal with small-circle concentrated directional data, association among these vectors usually occurs along small-circles with independent vertical errors. For example, when a 3D object is modeled by skeletal representations, as described in more detail in Section 2.6 and visualized in Fig. 8, a deformation of the object is measured by the movements of directional vectors on $\mathbb{S}^2$. When a single rotational deformation (such as bending, twisting or rotation) occurs, all the directions move along small-circles with a common axis μ_0. In this situation, the longitudinal variations along the circles are dependent on each other because nearby spoke vectors are under the effect of similar deformations. (Examples of such longitudinal dependencies can be found in Section 2.6 as well as in Schulz et al. (2015).) Adding such a horizontal (or longitudinal) dependence to a multivariate S1 model requires a careful introduction and parametrization of the offdiagonal entries of $\mathbf{A}$ in (2.11). This is not straightforward, and we leave it for future work. On the other hand, it is feasible to extend the S2 model by generalizing the "vMF part" of $\mathbf{y}$, the last term in the exponent of (2.12), to a Fisher-Bingham type.

To this end, we introduce a parameter matrix $\mathbf{B}$ to model general quadratics in $\mathrm{vec}(\mathbf{y})$. This allows to write the densities for a general multivariate small-sphere distribution of the second kind (GMS2) as follows:

$$\begin{aligned} & g_{\mathrm{GMS2}}(\mathbf{s}, \mathbf{y}; H, \mathbf{K}_0, \mathbf{M}, \mathbf{B}) \\ &= \frac{1}{T_1(H, \mathbf{K}_0) T_2(\mathbf{M}, \mathbf{B})} \exp\left\{H^\top \mathbf{s} - \mathbf{s}^\top \mathbf{K}_0 \mathbf{s} + \mathbf{M}^\top \mathrm{vec}(\mathbf{y}) + \mathrm{vec}(\mathbf{y})^\top \mathbf{B}\, \mathrm{vec}(\mathbf{y})\right\} \end{aligned} \tag{2.13}$$

where $H, \mathbf{K}_0$ and $\mathbf{M}$ as defined in (2.12), and $T_1(H, \mathbf{K}_0)$ and $T_2(\mathbf{M}, \mathbf{B})$ are normalizing constants. We set $\mathbf{B} = (B_{kl})_{k,l=1}^K$, $B_{kl} = (b_{i,j}^{(k,l)})_{i,j=1}^{p-1}$, as a block matrix with vanishing blocks $B_{kk} = 0$ on the diagonal. The submatrix B_{kl} models the horizontal association between y_k

and y_l. The fact that $z^\top \mathbf{B} z = z^\top \mathbf{B}^\top z = \frac{1}{2} z^\top (\mathbf{B} + \mathbf{B}^\top) z$ for any vector $z \in \mathbb{R}^{(p-1)K}$ allows us to assume without loss of generality that $\mathbf{B}$ is symmetric.

An MS2 distribution on $(\mathbb{S}^2)^K$. As a viable submodel for the practically important case $p = 3$, we propose to use a special form for the offdiagonal blocks B_{kl} of $\mathbf{B}$. In particular, with λ_{kl} representing the degrees of association between y_k and y_l, we set

$$\begin{aligned} B^{kl} &= 2 \begin{pmatrix} \widetilde{\mu_k} & \widetilde{\mu_k}^\perp \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \lambda_{kl} \end{pmatrix} \begin{pmatrix} \widetilde{\mu_l} & \widetilde{\mu_l}^\perp \end{pmatrix}^\top \\ &= 2\lambda_{kl}\, \widetilde{\mu_k}^\perp (\widetilde{\mu_l}^\perp)^\top, \end{aligned} \tag{2.14}$$

where $\begin{pmatrix} \widetilde{\mu_k} & \widetilde{\mu_k}^\perp \end{pmatrix}$ is the rotation matrix given by setting

$$\widetilde{\mu_k}^\perp = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \widetilde{\mu_k}.$$

The density (2.13) with the above parametrization of $\mathbf{B}$ will be referred to as a multivariate S2 distribution (MS2) for data on $(\mathbb{S}^2)^K$; its angular representation will be derived in (2.15) below.

Our choice of the simple parametrization (2.14) does not restrict the modeling capability of the general model (2.13), and has some advantages in parameter interpretations and also in estimation. To see this, we resort to use an angular representation for $\mathbf{y}$ (available to this $p = 3$ case). For each k, define ϕ_k and ζ_k such that $y_k = (\cos\phi_k, \sin\phi_k)^\top$ and $\widetilde{\mu_k} = (\cos\zeta_k, \sin\zeta_k)^\top$. Accordingly, the inner products appearing in (2.13) can be expressed as

$$\widetilde{\mu_k}^\top y_k = \cos(\phi_k - \zeta_k), \quad (\widetilde{\mu_k}^\perp)^\top y_k = \sin(\phi_k - \zeta_k).$$

Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)^\top$, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_K)^\top$, $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_K)^\top$,

$$\begin{aligned} c(\boldsymbol{\phi}, \boldsymbol{\zeta}) &= (\cos(\phi_1 - \zeta_1), \ldots, \cos(\phi_K - \zeta_K))^\top, \\ s(\boldsymbol{\phi}, \boldsymbol{\zeta}) &= (\sin(\phi_1 - \zeta_1), \ldots, \sin(\phi_K - \zeta_K))^\top, \end{aligned}$$

and $\mathbf{\Lambda} = (\lambda_{kl})_{k,l=1}^{K}$ where $\lambda_{kl} (= \lambda_{lk})$ for $k \neq l$ is the association parameter used in (2.14), and λ_{kk} is set to zero. The density of the MS2 distribution, in terms of $(\mathbf{s}, \boldsymbol{\phi})$, is then

$$\begin{aligned} & g_{\mathrm{MS2}}(\mathbf{s}, \boldsymbol{\phi}; H, \mathbf{K}_0, \boldsymbol{\kappa}, \boldsymbol{\zeta}, \mathbf{\Lambda}) \\ &= \frac{1}{T_1(H, \mathbf{K}_0) T_3(\boldsymbol{\kappa}, \mathbf{\Lambda})} \exp\left\{ H^\top \mathbf{s} - \mathbf{s}^\top \mathbf{K}_0 \mathbf{s} + \boldsymbol{\kappa}^\top c(\boldsymbol{\phi}, \boldsymbol{\zeta}) + \frac{1}{2} s(\boldsymbol{\phi}, \boldsymbol{\zeta})^\top \mathbf{\Lambda} s(\boldsymbol{\phi}, \boldsymbol{\zeta}) \right) \end{aligned} \tag{2.15}$$

From (2.15), it can be easily seen that the "horizontal angles" $\boldsymbol{\phi}$ follow the *multivariate von Mises* distribution (Mardia et al., 2008) and are independent of the vertical component $\mathbf{s}$. As we will see later in Section 2.3.2, this facilitates estimation for the MS2 distributions. Moreover, since

$$\begin{aligned} & \boldsymbol{\kappa}^\top c(\boldsymbol{\phi}, \boldsymbol{\zeta}) + \frac{1}{2} s(\boldsymbol{\phi}, \boldsymbol{\zeta})^\top \mathbf{\Lambda} s(\boldsymbol{\phi}, \boldsymbol{\zeta}) \\ &= \sum_{k=1}^{K} \kappa_k \left(1 - \frac{(\phi_k - \zeta_k)^2}{2} \right) + \frac{1}{2} \sum_{k=1}^{K} \sum_{k \neq l = 1}^{K} \left(\lambda_{kl} (\phi_k - \zeta_k)(\phi_l - \zeta_l) \right) + o(\|\boldsymbol{\phi} - \boldsymbol{\zeta}\|^2) \end{aligned} \tag{2.16}$$

for large enough concentrations, $\boldsymbol{\phi}$ is approximately multivariate normal with mean $\boldsymbol{\zeta}$ and precision matrix Σ^{-1}, where $(\Sigma^{-1})_{kk} = \kappa_k$ and $(\Sigma^{-1})_{kl} = -\lambda_{kl}$ for $1 \leq k \neq l \leq K$. These parameters are naturally interpreted as partial variances and correlations. This interpretation of the parameters as entries of a precision matrix is most immediate under the MS2, but is not under the general case.

2.2.4 Random data generation

Generating pseudo-random samples from the S1 and S2 distributions are important in simulations and in developments of computer-intensive inference procedures.

For simulation of the S1 (2.5) and iMS1 (2.11) distribution, the fact that each marginal distribution of the iMS1 is a special case of the Fisher-Bingham is handy. Thereby, one can use the Gibbs sampling procedure developed for generating Fisher-Bingham-variate samples (Hoff, 2009).

For simulation of the S2 (2.6), iMS2 (2.12), and MS2 (2.15) distribution, we take advantage of the independence of $\mathbf{s}$ and $\mathbf{y}$. As we assume vertical independence (i.e., $s_1, \ldots, s_K$ are independent), each s_k can be sampled separately. Therefore, sampling from the MS2

distribution amounts to independently drawing samples from a truncated normal distribution (for s_k) and from a multivariate von Mises distribution (for $\mathbf{y}$). Specifically, to sample $\mathbf{x} = (x_1, \ldots, x_K)$ from MS2$(\mu_0, \boldsymbol{\mu}, \boldsymbol{\kappa}_0, \boldsymbol{\kappa}, \boldsymbol{\Lambda})$, the following procedure can be used.

Step 1. For each k, sample s_k from the truncated normal distribution with mean ν_k and variance $1/(2\kappa_{0k})$, truncated to the interval (-1,1).

Step 2. For the S2 or iMS2 model, sample each $y_k \in \mathbb{S}^{p-2}$ in $\mathbf{y} = (y_1, \ldots, y_K)$ independently from the von Mises distribution with mean $(1, 0, \ldots, 0)$ and concentration κ_k; for the MS2 distribution (when $p = 3$), sample the K-tuple $\mathbf{y} \in (\mathbb{S}^1)^K$ directly from the multivariate von-Mises distribution with mean $(1, 0)$ and precision parameters $\boldsymbol{\kappa}$ and $\boldsymbol{\Lambda}$.

Step 3. For each k, let E_k be a $p \times p$ orthogonal matrix with $(\mu_0, P_{\mu_0}\mu_k/\|P_{\mu_0}\mu_k\|)$ being the first two column vectors. Set $x_k = E_k^\top \left(s_k, (1 - s_k^2)^{-1/2} y_k\right)$.

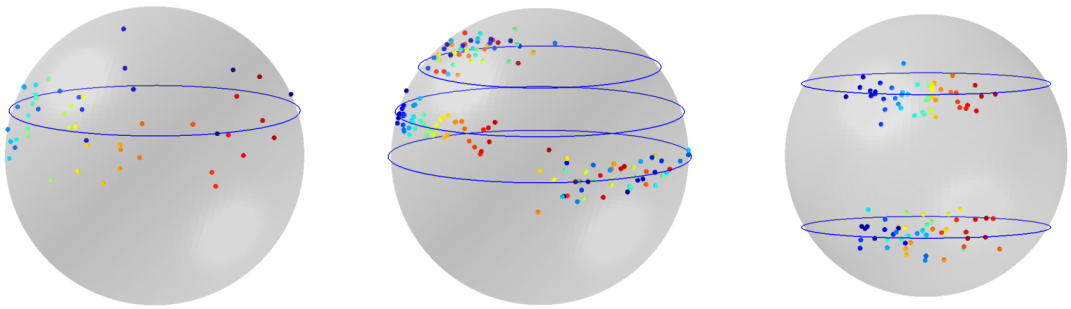
In our experiments, sampling from the S2 and MS2 distributions is much faster than from the S1. In particular, when the dimension p or the concentration level is high, the Markov chain simulations for the S1 appear to be sluggish. Some examples of random samples from the S2, iMS2 and MS2 distributions are shown in Fig. 5. The small-circles $C(\mu_0, \nu_k)$ are also overlaid in the figure. Notably, the MS2 sample in the rightmost panel clearly shows a horizontal dependence.

2.3 MAXIMUM LIKELIHOOD ESTIMATION

The algorithms developed below converge quickly to a local maximum of the likelihood function in all of the data situations of this paper. In Section 2.8.2, we give an example of our new quick algorithm for the maximum likelihood estimation of S1.

2.3.1 Estimation for S1 and iMS1 models

The standard way to estimate parameters of the S1 is to use the maximum likelihood estimates (MLE). However, it does not seem possible to obtain explicit expressions of the MLE, partly due to having no closed-form expression of the normalizing constant (2.5). We propose



(a) S2 on $\mathbb{S}^2$ (b) iMS2 on $(\mathbb{S}^2)^3$ (c) MS2 on $(\mathbb{S}^2)^2$

Figure 5: Random samples from the S2, iMS2 and MS2 distributions. Same colors represent same simulations. (a) low concentrations ($\kappa_0 = 10$, $\kappa = 1$). (b) independent directions with high concentrations ($\kappa_{0k} = 100$, $\kappa_k = 10$, $k = 1, 2, 3$). (c) horizontally dependent directions with high concentrations ($\kappa_{0k} = 50$, $\kappa_k = 30$) and high dependence ($\lambda_{12} = 24$).

to approximate the normalizing constant, and numerically obtain the MLE. Our procedure naturally extends to estimation for the iMS1 distribution, which will also be discussed.

As a preparation, we first describe an approximation of the normalizing constant, following Kume and Wood (2005). The exact calculation of the normalizing constant of Fisher-Bingham distribution, including the S1, is possible (Kume and Sei, 2018), but we chose to use the approximation as the method of Kume and Sei (2018) is computationally too heavy for our application in mind.

The normalizing constant of the S1 has an alternative expression, as shown in the following.

Proposition 2. *For any $h > 0$, let $\xi = (\frac{\nu(2\kappa_0+\kappa_1)}{2(\kappa_0+h)}, \frac{\kappa_1\sqrt{1-\nu^2}}{2h}, 0, \ldots, 0)^\top \in \mathbb{R}^p$ and let Ψ be the $p \times p$ diagonal matrix with diagonal elements $(\kappa_0 + h, h, \ldots, h)$. Moreover, let $g(r)$ $(r > 0)$ be the probability density function of $R = Z^\top Z$, where $Z \sim N_p(\xi, \frac{1}{2}\Psi^{-1})$. Then the normalizing constant $a(\kappa_0, \kappa, \nu)$ of (2.5) is*

$$a(\kappa_0, \kappa_1, \nu) = 2\pi^{p/2}|\Psi|^{-1/2} g(1) \exp\left(\xi^T \Psi \xi + h - \kappa_0 \nu^2\right). \tag{2.17}$$

In Proposition 2, the function g is the density of a linear combination of independent noncentral χ^2_1 random variables. Following Kume and Wood (2005), we use saddle-point density approximations in the numerical computation of $g(1)$. First, note that the derivatives of the cumulant generating function, $K_g(t) = \log \int_0^\infty e^{tr} g(r)dr$, associated with the density g have closed-form expressions. Denoting by $K_g^{(j)}(t)$ the jth derivative of $K_g(t)$, for $j = 1, \ldots, 4$, we get

$$K_g^{(j)}(t) = \frac{(j-1)!}{2}\left(\frac{1}{(\kappa_0 + h - t)^j} + \frac{p-1}{(h-t)^j}\right) + \frac{j!}{4}\left(\frac{\nu^2(2\kappa_0+\kappa_1)^2}{(\kappa_0+h-t)^{j+1}} + \frac{\kappa_1^2(1-\nu^2)}{(h-t)^{j+1}}\right).$$

Let $\hat{t}$ be the unique solution in $(-\infty, h)$ of the saddle-point equation $K_g^{(1)}(t) = 1$, which can be easily evaluated by using, e.g., a bisection method. Then a saddle-point density approximation of $g(1)$ is

$$\hat{g}(1) = (2\pi K_g^{(2)}(\hat{t}))^{-1/2} \exp(K_g(\hat{t}) - \hat{t} + T), \tag{2.18}$$

where $T = K_g^{(4)}(\hat{t})/\{8(K_g^{(2)}(\hat{t}))^2\} - 5(K_g^{(3)}(\hat{t}))^2/\{24(K_g^{(2)}(\hat{t}))^3\}$. In the following, we approximate the value of $a(\kappa_0, \kappa_1, \nu)$ by $\hat{a}(\kappa_0, \kappa_1, \nu)$ obtained by plugging (2.18) in place of $g(1)$ in (2.17).

We are now ready to describe our estimation procedure. Suppose $x_1, \ldots, x_n$ is a sample from $\mathrm{S1}(\mu_0, \mu_1, \kappa_0, \kappa_1)$ and let $\ell_n(\mu_0, \mu_1, \kappa_0, \kappa_1, \nu)$ be the log-likelihood.

Suppose for now that $\nu \in [0, 1)$ is fixed. Then the MLE of μ_0 and μ_1 can be efficiently estimated. In particular, maximizing the likelihood function with respect to μ_0 is equivalent to minimizing $\frac{1}{n}\sum_{i=1}^n (\mu_0^\top x_i - \nu)^2$ subject to the constraint $\mu_0^\top \mu_0 = 1$. With a Lagrangian multiplier λ using matrix notation, we solve

$$\min_{\mu_0 \in \mathbb{S}^{p-1}} \left[\frac{1}{n}\|\mathbb{X}^\top \mu_0 - \nu 1_n\|^2 - \lambda(\mu_0^\top \mu_0 - 1)\right], \tag{2.19}$$

where $\mathbb{X}$ is the $p \times n$ matrix whose ith column is x_i, yielding the necessary condition $S\mu_0 - \nu\bar{x} - \lambda\mu_0 = 0$, where $S = \mathbb{X}\mathbb{X}^\top/n$, $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$. For a fixed Lagrangian multiplier λ, the solution is $\hat{\mu}_0 = \nu(S - \lambda I_p)^{-1}\bar{x}$, provided that S is of full rank. The constraint $\mu_0^\top \mu_0 = 1$ makes us find a root λ of $\nu^2 \bar{x}^\top (S - \lambda I_p)^{-2}\bar{x} - 1$. The root $\hat{\lambda}$ is found by a bisection search in

the range $[-\nu^2\bar{x}^\top\bar{x}, \lambda_S]$, where $\lambda_S > 0$ is the smallest eigenvalue of S (Browne, 1967). The solution to (2.19) is then

$$\hat{\mu}_0 = \nu(S - \hat{\lambda}I_p)^{-1}\bar{x}. \tag{2.20}$$

If $\nu = 0$, then $\hat{\mu}_0$ is the eigenvector of S corresponding to the smallest eigenvalue.

Now, with ν and $\hat{\mu}_0$ (2.20) given, maximizing the likelihood with respect to μ_1 is equivalent to maximizing $\frac{1}{n}\sum_{i=1}^n \mu_1^\top x_i$ subject to the constraints $\mu_0^\top\mu_1 = \nu$ and $\mu_1^\top\mu_1 = 1$. It can be shown that the MLE of μ_1 is a linear combination of μ_0 and $\bar{x}/\|\bar{x}\|$ where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$. Let $\bar{x}^* = P_{\mu_0}\bar{x}/\|P_{\mu_0}\bar{x}\|$, then the MLE of μ_1 given ν and $\hat{\mu}_0$ is

$$\hat{\mu_1} = \nu\hat{\mu}_0 + \sqrt{1-\nu^2}\bar{x}^*. \tag{2.21}$$

Thus, the MLE of $(\kappa_0, \kappa_1, \nu)$ is

$$(\hat{\kappa}_0, \hat{\kappa}_1, \hat{\nu}) = \arg\max_{\kappa_0,\kappa_1,\nu} \ell_n(\hat{\mu}_0, \hat{\mu}_1, \kappa_0, \kappa_1, \nu),$$

which is solved by a standard optimization package, and the MLE of (μ_0, μ_1) is given by (2.20) and (2.21) with ν replaced by $\hat{\nu}$. This procedure, beginning with an initial value for $\hat{\nu}$, is iterated until convergence.

Let us now describe an extension of the above algorithm to the iMS1 model. Suppose $(x_{i1}, \ldots, x_{iK}) \in (S^{p-1})^K$ for $i = 1, \ldots, n$ is a sample from an iMS1 model, where each marginal distribution is $\mathrm{S1}(\mu_0, \mu_k, \kappa_{0k}, \kappa_k)$. While the last step above can be applied to estimate $\kappa_{0k}, \kappa_k, \nu_k$, for $k = 1, \ldots, K$, given μ_0 and μ_k's, we replace (2.19) with

$$\min_{\mu_0\in\mathbb{S}^{p-1}} \left[\frac{1}{n}\sum_{k=1}^K (\kappa_{0k}\|\mathbb{X}_k^\top\mu_0 - \nu_k 1_n\|^2) - \lambda(\mu_0^\top\mu_0 - 1)\right],$$

where the marginal $p \times n$ observation matrix $\mathbb{X}_k$ has the columns x_{ik} $(i = 1, \ldots, n)$. This is solved with the obvious analog to (2.20). For μ_k's, the above solution (2.21) can be applied for each k separately.

2.3.2 Estimation for S2, iMS2 and MS2

The S2 model and its extensions have the convenient property that the horizontal components are independent of the vertical ones. To take advantage of this, suppose for now that μ_0 is known. This allows us to decompose an observation x into two independent random variables (s, y), which in turn leads to an easy estimation of the remaining parameters $\eta := (\mu, \kappa_0, \kappa_1)$. Thus, the estimation strategy proceeds in two nested steps. Let $\ell_n(\mu_0, \eta)$ be the log-likelihood function given a sample $x_1, \ldots, x_n$ from S2(μ, η). In the *outer step*, we update μ_0 to maximize the likelihood, i.e.,

$$\hat{\mu}_0 = \operatorname*{argmin}_{\mu_0} \ell_n(\mu_0, \hat{\eta}_{\mu_0}), \tag{2.22}$$

where evaluating

$$\hat{\eta}_{\mu_0} = \operatorname*{argmin}_{\eta} \ell_n(\mu_0, \eta) \tag{2.23}$$

for a fixed μ_0 is the *inner step*. It is straightforward to see that the MLE of (μ_0, η) is given by $(\hat{\mu}_0, \hat{\eta}_{\hat{\mu}_0})$.

In the following, we discuss in detail the inner step (2.23) of maximizing $\ell_{\mu_0}(\eta) := \ell_n(\mu_0, \eta)$ for the iMS2 model (2.12) and for the MS2 model (2.15), while we resort to a standard optimization package for solving (2.22).

Independent multivariate S2 (iMS2). Suppose $(x_{i1}, \ldots, x_{iK}) \in (S^{p-1})^K$ for $i = 1, \ldots, n$ is a sample from an iMS2, where each marginal distribution is S2$(\mu_0, \mu_k, \kappa_{0k}, \kappa_k)$. For a given μ_0, the joint density can be written in terms of $(\mathbf{s}_i, \boldsymbol{\phi}_i)$ as done in (2.15), but with $\boldsymbol{\Lambda} = 0$. Furthermore, by the definition of H and $\mathbf{K}_0$ we used in (2.15), we can write

$$H^\top \mathbf{s}_i - \mathbf{s}_i^\top \mathbf{K}_0 \mathbf{s}_i = -\sum_{k=1}^{K} \kappa_{0k}(s_{ik} - \nu_k)^2 + \sum_{k=1}^{K} \kappa_{0k}\nu_k^2,$$

and hence

$$\log\left[T_1(H, \mathbf{K}_0) T_3(\boldsymbol{\kappa}, 0)\right] = \sum_{k=1}^{K} \left[\log b(\kappa_{0k}, \kappa_k, \nu_k) + \kappa_{0k}\nu_k^2\right].$$

Note that the normalizing constant $b(\kappa_0, \kappa_1, \nu)$ satisfies

$$\begin{aligned} b(\kappa_0, \kappa_1, \nu) &= \int_0^{2\pi} e^{\kappa_1 \cos\phi} d\phi \int_{-1}^{1} e^{-\kappa_0(s-\nu)^2} ds \\ &= (2\pi)^{3/2}(2\kappa_0)^{-1/2} \mathcal{I}_0(\kappa_1) \left[\Phi((1-\nu)\sqrt{2\kappa_0}) - \Phi(-(1+\nu)\sqrt{2\kappa_0})\right], \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Finally, the log-likelihood function (given μ_0) is, for $\boldsymbol{\kappa}_0 = (\kappa_{01}, \ldots, \kappa_{0K})^\top$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_K)^\top$,

$$\ell_{\mu_0}(\boldsymbol{\nu}, \boldsymbol{\zeta}, \boldsymbol{\kappa}_0, \boldsymbol{\kappa}; \{\mathbf{s}_i, \boldsymbol{\phi}_i\}_{i=1}^n) = \ell_{\mu_0}^{(1)}(\boldsymbol{\nu}, \boldsymbol{\kappa}_0) + \ell_{\mu_0}^{(2)}(\boldsymbol{\zeta}, \boldsymbol{\kappa}), \tag{2.24}$$

where

$$\begin{aligned} \ell_{\mu_0}^{(1)}(\boldsymbol{\nu}, \boldsymbol{\kappa}_0) &= -\sum_{k=1}^{K} \left[\kappa_{0k} \sum_{i=1}^{n} (s_{ik} - \nu_k)^2 - \frac{n}{2} \log(2\kappa_{0k}) + \frac{n}{2} \log(2\pi) \right. \\ &\quad \left. + n \log\left(\Phi\left((1-\nu_k)\sqrt{2\kappa_{0k}} \right) - \Phi\left(-(1+\nu_k)\sqrt{2\kappa_{0k}} \right) \right) \right], \end{aligned} \tag{2.25}$$

$$\ell_{\mu_0}^{(2)}(\boldsymbol{\zeta}, \boldsymbol{\kappa}) = \sum_{k=1}^{K} \left[\kappa_k \sum_{i=1}^{n} \cos(\phi_{ik} - \zeta_k) - n \log \mathcal{I}_0(\kappa_k) - n \log(2\pi) \right]. \tag{2.26}$$

Therefore, the optimization for the inner step (2.23) is equivalent to simultaneously solving $2K$ subproblems.

Each of the K subproblems of (2.25) is equivalent to obtaining the MLE of a truncated normal distribution $\mathrm{trN}(\nu_k, (2\kappa_{0k})^{-1/2}; (-1, 1))$ based on the observations s_{ik} $(i = 1, \ldots, n)$. Similarly, each of the K subproblems of (2.26) amounts to obtaining the MLE of a von Mises distribution with mean ζ_k and concentration κ_k from the sample ϕ_{ik} $(i = 1, \ldots, n)$. The MLEs of the truncated normal are numerically computed, and we use the method of Banerjee et al. (2005) to obtain approximations of the MLEs of the von Mises.

MS2. Under the MS2 model (2.15) with a dependence structure on $\boldsymbol{\phi}_i$, a decomposition $\ell_{\mu_0}(\boldsymbol{\nu}, \boldsymbol{\zeta}, \boldsymbol{\kappa}_0, \boldsymbol{\kappa}, \boldsymbol{\Lambda}) = \ell_{\mu_0}^{(1)}(\boldsymbol{\nu}, \boldsymbol{\kappa}_0) + \ell_{\mu_0}^{(2)}(\boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\Lambda})$, similar to (2.24), is valid, where (2.26) is replaced by

$$\ell_{\mu_0}^{(2)}(\boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\Lambda}) = -\sum_{i=1}^{n} \left[\boldsymbol{\kappa}^\top c(\boldsymbol{\phi}_i, \boldsymbol{\zeta}) + \frac{1}{2} s(\boldsymbol{\phi}_i, \boldsymbol{\zeta})^\top \boldsymbol{\Lambda} s(\boldsymbol{\phi}_i, \boldsymbol{\zeta}) - \log T_3(\boldsymbol{\kappa}, \boldsymbol{\Lambda}) \right]. \tag{2.27}$$

Maximizing (2.27) is equivalent to computing the MLE of the multivariate von Mises distribution (Mardia et al., 2008). We either use maximum pseudo-likelihood estimate as discussed in Mardia et al. (2008) or moment estimates, yielding

$$\hat{\zeta}_k = \frac{1}{n} \sum_{i=1}^{n} \phi_{ik} / \| \frac{1}{n} \sum_{i=1}^{n} \phi_{ik} \|, \quad \hat{\kappa}_k = \bar{S}_{kk}^{-1}, \quad \hat{\lambda}_{(kl)} = \bar{S}_{kl}^{-1} \quad (k \neq l),$$

where $\bar{S} = (\bar{S}_{kl})$ and $\bar{S}_{kl} = \frac{1}{n}\sum_{i=1}^{n} \sin(\phi_{ik} - \hat{\zeta}_k)\sin(\phi_{il} - \hat{\zeta}_l)$ for $k, l = 1, \ldots, K$. These estimates coincide with the MLEs when $K = 2$. For larger $K > 3$, the accuracy of the moment estimates deteriorates, but evaluating MLEs or a maximum pseudo-likelihood estimator becomes computationally highly expensive.

2.4 TESTING HYPOTHESES

It is of interest to infer on the parameters of our models. While we adopt the well-known likelihood ratio test, we emphasize that the proposed model enables us to test several important hypotheses, which have been of interest to some researchers, and that our estimation procedure can be easily adapted to compute maximized likelihood under the null parameter space Θ_0. Recall that the parameter space for the iMS1 and iMS2 models is given by $\Theta_{\text{ind}} = \mathbb{S}^{p-1} \times (\mathbb{S}^{p-1})^K \times (\mathbb{R}_+)^K \times (\mathbb{R}_+)^K$ for $\theta_{\text{ind}} = (\mu_0, \boldsymbol{\mu}, \boldsymbol{\kappa}_0, \boldsymbol{\kappa})$. For conciseness, we describe our testing procedure using the MS2 distribution in dimension $p = 3$, whose parameter space is $\Theta = (\Theta_{\text{ind}} \times (\mathbb{R})^{K(K-1)/2})$ for $\theta = (\theta_{\text{ind}}, \boldsymbol{\Lambda})$. For some Θ_0 that dictates a null hypothesis H_0 and satisfies $\Theta_0 \subset \Theta$, we denote the maximized log-likelihood under Θ_0 by ℓ_0, and the maximized log-likelihood under Θ by ℓ_1. Our test statistic is $W_n := -2(\ell_0 - \ell_1)$ and H_0 is rejected for large enough W_n.

We list a few null hypotheses of practical interest, with the alternative being the full MS2 distribution. In all three cases below, the alternative hypothesis is $H_1 : \theta \in \Theta \setminus \Theta_0$.

1. **Test of association**. H_0: $\boldsymbol{\Lambda} = \mathbf{0}$, i.e., there is no horizontal dependence.

2. **Test of axis**. H_0: $\mu_0 = \mu_0^*$, i.e., the common axis of rotation is μ_0^*.

3. **Test of great-sphere**. H_0: $\boldsymbol{\nu} = 0$, i.e., the underlying subsphere is indeed a great subsphere.

While the test of association is only available under the MS2 model ($p = 3$), Hypotheses 2 and 3 can also be tested using S1, S2, iMS1 or iMS2 models in any dimension $p \geq 3$.

To validate the use of small-sphere distributions, in any dimension $p \geq 3$, the following hypotheses can be tested.

4. **Test for von Mises-Fisher distribution**. H_0: $\kappa_0 = 0$, i.e., there is no "small-circle

feature."

5. **Test for Bingham-Mardia distribution**. H_0: $\kappa = 0$, i.e., there is no unique mode.

For each hypothesis, computing the test statistic W_n requires maximizing the likelihood on Θ_0 (or to compute ℓ_0). This is easily achieved by modifying the iterative algorithms in Section 2.3. For example, for the test of association, computing ℓ_0 and ℓ_1 amounts to obtaining the MLEs under the iMS2 and MS2 models, respectively; for Hypothesis 2 (test of axis), where μ_0 is given, one only needs to solve (2.23) once. Other cases of restricted MLEs can be easily obtained. In Section 2.8.3, we confirm that the test statistic W_n using our algorithms under the null hypotheses above is empirically nearly chi-square distributed for sample size $n = 30$. In Section 2.5.3 and in Section 2.8.3, empirical powers of the proposed test procedures are reported for several important alternatives.

2.5 NUMERICAL STUDIES

We demonstrate the performances of small-circle fitting in Section 2.5.1, the ability of the MS2 of modeling the horizontal dependence in Section 2.5.2 and a testing procedure to prevent overfitting in Section 2.5.3.

2.5.1 Estimation of small-circles

The performance of our estimators in fitting the underlying small-spheres $\mathcal{C}(\mu_0, \nu)$ is numerically compared with those of competing estimators obtained from assuming the Bingham-Mardia (BM) distribution and the least-square methods of Schulz et al. (2015). The BM distribution has originally been defined only for data on $\mathbb{S}^2$, but we use a natural extension given by a special case of the iMS1. Thus, "BM estimates" refer to the estimates of the iMS1 model with the restriction $\boldsymbol{\kappa} = 0$. The estimates of Schulz et al. are obtained by minimizing the sum of squared angular distances from observations to $\mathcal{C}(\hat{\mu}_0, \hat{\nu})$, which will be referred to as a "least-squares (LS)" method.

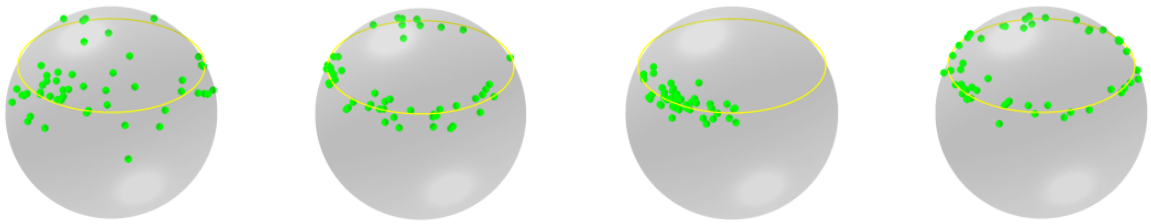
We first consider four univariate S2 models to simulate data concentrated on a single

small circle in order to benchmark against existing BM and LS. The directional parameters (μ_0, μ_1) are set to satisfy $\nu = 0.5$. We use $(\kappa_0, \kappa_1) = (10, 1), (100, 1), (100, 10)$ to represent various data situations. Random samples from these three settings are shown in Fig. 6(a)–(c). We also consider the BM model as a special case of the S2 distributions (by setting $\kappa_1 = 0$); a sample from the BM distribution is shown in Fig. 6(d).

The small-circle estimation performances of the S1, S2, BM, and LS estimates are measured by an angular product error (in degrees), defined as

$$L\left((\mu_0, \nu), (\hat{\mu}_0, \hat{\nu})\right) = \left(\text{Angle}(\hat{\mu}_0, \mu_0)^2 + \left(\frac{180}{\pi}(\arccos\hat{\nu} - \arccos\nu)\right)^2\right)^{1/2}. \tag{2.28}$$

Table 2 displays the means and standard deviations of $L\left((\mu_0, \nu), (\hat{\mu}_0, \hat{\nu})\right)$ from 100 repetitions for each of the four methods, fitted to random samples of size 50 from each of the settings, labeled (a)–(d). Since horizontal/vertical dependence is not of concern, we expect that S1, S2 and LS perform similarly and that they strongly outperform BM when the small circle features a distinct mode (c). This is indeed the case. Remarkably, in the high vertical noise case (a), the standard deviations of S1 and S2 are considerably lower than that of BM and still notably smaller than that of LS.



(a) $\kappa_0 = 10, \kappa_1 = 1$ (b) $\kappa_0 = 100, \kappa_1 = 1$ (c) $\kappa_0 = 100, \kappa_1 = 10$ (d) $\kappa_0 = 100, \kappa_1 = 0$

Figure 6: Random samples of size $n = 50$ from the S2 model on $\mathbb{S}^2$ used in our simulations. Small-circle estimation performances are reported in Table 2.

Method	(a)	(b)	(c)	(d)
S1	5.81(3.17)	1.46(0.72)	15.56(11.68)	**1.35**(0.72)
S2	**5.24**(3.14)	**1.44**(0.70)	**14.36**(11.91)	1.36(0.72)
BM	8.37(6.43)	1.47(0.70)	17.00(15.68)	**1.35**(0.72)
LS	5.83(3.74)	1.45(0.69)	15.27(14.16)	1.36(0.73)

Table 2: Small-circle estimation performances for univariate data on $\mathbb{S}^2$ from Fig. 6. Means (standard deviations) of the angular product errors in degrees (2.28) are shown.

Next, to show the performance of our multivariate models, we consider six bivariate MS2 models. The directional parameters $(\mu_0, \boldsymbol{\mu})$ were set to satisfy $\boldsymbol{\nu} = (0.5, -.3)$, and the concentration parameters were chosen to mimic the concentrations of the univariate models, described above. For Cases (a)–(c), we set $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (10, 1, 0), (100, 1, 0), (100, 10, 0)$, for $j = 1, 2$, so that the models are indeed the iMS2. For the latter three cases (d)–(f), we set $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (10, 2, 1.5), (100, 2, 1.5), (100, 20, 15)$, $j = 1, 2$, to make their vertical and horizontal dispersions be similar to the iMS2 counterparts. By setting $\lambda_{12} > 0$, the random bivariate directions are positively associated. (Examples of random samples from these settings can be found in Fig. 15.) The small-circle estimation performance of the iMS1, iMS2, MS2, BM and LS estimates is measured by the canonical multivariate extension of the angular product error (2.28). Table 3 collects the means and standard deviations of the angular product errors from 100 repetitions with the sample size $n = 50$. Now BM clearly performs worst, also in low horizontal concentration scenarios (a), (b), (d) and (e). Since BM is able to model vertical but not horizontal concentration, somewhat unexpected, BM performs considerably worse under additional high vertical noise (a) and (d), as compared to (b) and (e). The same phenomenon, but more subtly is also visible for LS. In case of high vertical concentration (all but (a) and (d)), LS performs comparable to the new parametric models. In particular in case of additional horizontal dependence with high concentration (f), it outperforms the independent parametric models and it is as good as MS2. In case of

	Independent			Dependent		
Method	(a)	(b)	(c)	(d)	(e)	(f)
iMS1	4.84(1.75)	**1.51(0.56)**	**3.90(3.61)**	6.21(2.73)	1.58(0.76)	4.91(5.82)
iMS2	4.49(1.80)	1.52(0.53)	4.30(2.46)	5.90(2.60)	1.58(0.75)	4.55(2.56)
MS2	**4.47(1.78)**	1.52(0.53)	4.26(2.45)	**5.81(2.67)**	**1.57(0.75)**	**4.50(2.48)**
BM	5.30(2.21)	1.54(0.54)	8.63(4.66)	13.87(15.42)	1.69(0.83)	9.26(5.25)
LS	4.76(1.85)	1.52(0.55)	4.30(2.35)	6.68(2.77)	1.59(0.71)	4.53(2.44)

Table 3: Small-circles estimation performances for bivariate data on $\mathbb{S}^2 \times \mathbb{S}^2$ from Fig. 15. Means (standard deviations) of the angular product errors in degrees (2.28) are shown.

low vertical concentration (a), however, MS2 is superior, MS2 is considerably superior under additional horizontal dependence (d).

We check robustness against model misspecification of the estimators by simulating data from a more general signal-plus-noise model (neither S1 nor S2). The performances of small-circle fitting of the proposed methods are comparable to that of the least-square estimator. Relevant simulation results and a detailed discussion can be found in Section 2.8.4.

2.5.2 Estimation of horizontal dependence

The ability of the MS2 to model the horizontal dependence is an important feature of the proposed distributions. We emphasize that the MS2 is the only method modeling such dependence, hence in this section we can only validate it against iMS2. Here, we empirically confirm that the MS2 estimates provide accurate measures of horizontal dependence, using Cases (c) and (f) in Section 2.5.1. For sample sizes $n = 50$ and 200, the concentration and association parameters were estimated under the assumption of MS2 (or iMS2), and Table 4 summarizes the estimation accuracy. In all cases, the MS2 model provides precise

		(c)			(f)		
n	Method	$\kappa_1 = 10$	$\kappa_2 = 10$	$\lambda_{12} = 0$	$\kappa_1 = 20$	$\kappa_2 = 20$	$\lambda_{12} = 15$
50	iMS2	10.23(2.48)	10.54(2.16)		11.73(2.19)	11.19(2.14)	
	MS2	10.48(2.54)	10.80(2.27)	−0.17(1.85)	22.63(5.06)	21.40(4.32)	16.82(4.22)
200	iMS2	10.31(1.08)	10.10(1.04)		11.00(1.05)	11.09(1.04)	
	MS2	10.35(1.10)	10.14(1.05)	−0.12(0.73)	20.38(2.17)	20.54(2.25)	15.41(2.06)

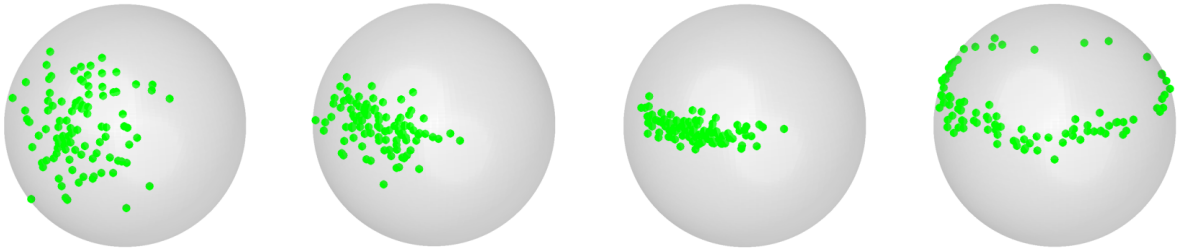
Table 4: Concentration and association parameter estimates for bivariate data on $\mathbb{S}^2 \times \mathbb{S}^2$ from Fig. 15. Means (standard deviations) of the estimates (from 100 repetitions). The column headings show the true parameters.

estimations of the horizontal dispersion and dependence; as the sample size increases, the mean squared error decreases. For Case (c), the underlying model is exactly iMS2, so the iMS2 estimates have smaller mean squared errors than the MS2 estimates. However, for Case (f), we notice that the iMS2 estimates of $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)$ become inferior. In fact, in case of existing horizontal dependence, i.e., when $\lambda_{12} \neq 0$, the concentration parameters $\boldsymbol{\kappa}$ in the misspecified iMS2 model do not correctly represent the concentrations as correctly represented by the MS2 model. This is so because the marginal distribution of ϕ_j, $j = 1, 2$, in (2.15) is not a von Mises distribution (Shing et al., 2002; Mardia et al., 2008).

2.5.3 Detecting overfitting in an isotropic case

When the data do not exhibit a strong tendency of a small-circle feature, the S1 and S2 distributions may overfit the data. For example, to a random sample from an isotropic vMF distribution as shown in Fig. 7(a), the S1 or S2 model fits an unnecessary small-circle $\mathcal{C}(\hat{\mu}_0, \hat{\nu})$. Indeed, a small-circle fit was observed in 83% of simulations of fitting the S1 model. Using the BM or LS results in a similar overfitting, where very small circles are erroneously fitted for 100% and 68% of the simulations, for the BM and LS, respectively.

This problem of overfitting has been known for a while and discussed in the context of dimension reduction of directional data. In particular, Jung et al. (2011, 2012) and



(a) vMF ($\kappa_1 = 10$) (b) $(\kappa_0, \kappa_1) = (20, 10)$ (c) $(\kappa_0, \kappa_1) = (100, 10)$ (d) $(\kappa_0, \kappa_1) = (100, 1)$

Figure 7: Degrees of the "small-circle feature." Shown are random samples from an isotropic distribution (case (a)), and the S2 distributions with increasing "small-circle concentrations" (cases (b)–(d)).

Eltzner et al. (2017) investigated the overfitting phenomenon for the least-square estimates and proposed some ad-hoc methods for adjustment. To prevent the overfitting, we point out that the testing procedure in Section 2.4 for the detection of isotropic distributions (Hypothesis 4) works well. To confirm this, we evaluated the empirical power of the test at the significance level $\alpha = 0.05$ for several alternatives. The power increases sharply as the distributions become more anisotropic; under the alternative distributions depicted in Fig. 7(b)–(d), the empirical powers are respectively $\hat{\beta} = 0.435, 1$ and 1, evaluated from 200 repetitions.

2.6 ANALYSIS OF S-REP DATA

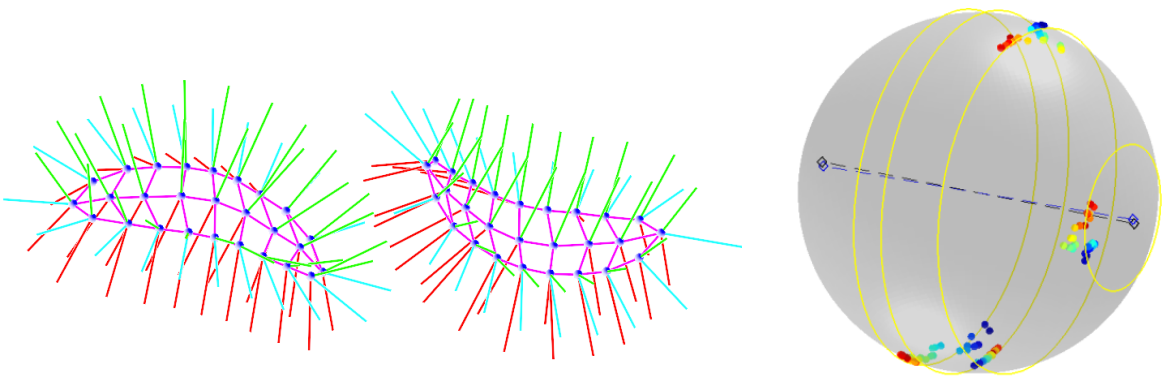
In this section, an application of the proposed distributions and test procedures to s-rep data is discussed.

2.6.1 Modeling rotationally-deformed ellipsoids via s-reps

Skeletal representations (s-reps) have been useful in mathematical modeling of anatomical objects (Siddiqi and Pizer, 2008). Roughly, an s-rep model for a 3-dimensional object consists of locations of a skeletal mesh (inside of the object) and spoke vectors (directions and lengths), connecting the skeletal mesh with the boundary of the object; examples are shown in the top left panel of Fig. 8. When the object is "rotationally deformed", Schulz et al. (2015) have shown that the directional vectors of an s-rep model approximately trace a set of concentric small-circles on $\mathbb{S}^2$, as shown in the top panels of the figure. Such rotational deformations (e.g., rotation, bending and twisting) of human anatomical objects have been observed in between and within shape variations of hippocampi and prostates (Joshi et al., 2002; Jung et al., 2011; Pizer et al., 2013). We demonstrate the use of the MS2 distribution in modeling (and fitting) a population of such objects via s-reps. Note that the sample space of an s-rep with K spokes is $(\mathbb{S}^2)^K \times \mathbb{R}_+^K \times (\mathbb{R}^3)^K$ (for direction, length, and location). In this work, we choose to analyze the spoke directions in $(\mathbb{S}^2)^K$ only, leaving a full-on analysis, accommodating the lengths and locations, to future work.

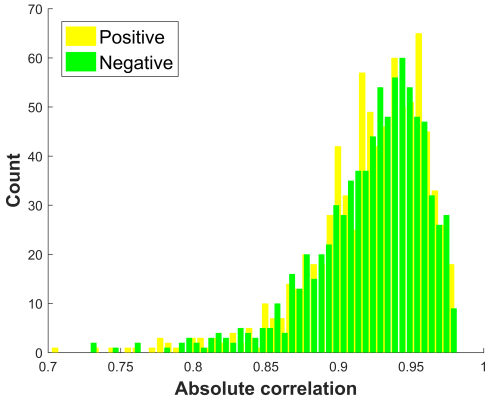
2.6.2 Data preparation

For our purpose of validating the use of the proposed distributions, we use an s-rep data set, fitted from $n = 30$ deformed ellipsoids; two samples from this data set are shown in Fig. 8(a). This data set was previously used in Schulz et al. (2015) as a simple experimental representation of real human organs. The data set was generated by "physically bending" a template ellipsoid about an axis $\mu_0^* = (0, 1, 0)$ by random angles drawn from a normal distribution with standard deviation 0.4 (radians). Each deformed ellipsoid is then recorded as a 3-dimensional binary image. To mimic the procedure of fitting s-reps from, for example, medical resonance imaging of a real patient, s-reps with 74 spokes were fitted to these binary images. (See Pizer et al. (2013) for details of the s-rep fitting.) As a preprocessing, we chose $K = 58$ spoke vectors, excluding the vectors with very small total variation.

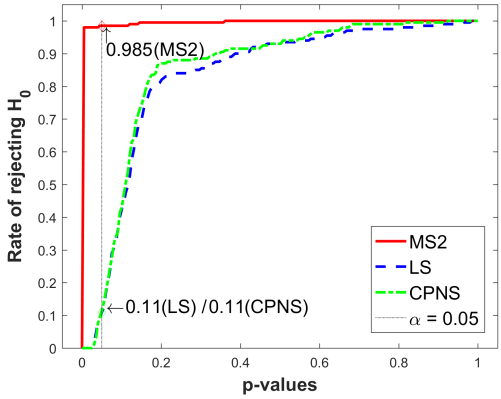


(a)

(b)



(c)



(d)

Figure 8: (a) Two s-rep models of randomly-bent ellipsoids. Skeletal mesh points (blue) due to grid lines (purple) with spokes (green, red and cyan). (b) Directions-circles plot: Graphical display of MS2 parameter estimates (small-circles (yellow), $\hat{\mu}_0$ (blue dashed axis) compared with μ_0^* (black dashed)) laid over the data where same colors correspond to same simulations. (c) Histogram of estimated "horizontal" correlation coefficients. (d) Empirical distributions of p-values from horizontal-dependence tests. See text for details.

2.6.3 Inference on the bending axis

Fitting the iMS1 distribution, we obtained the axis estimate $\hat{\mu}_0^{(\text{iMS1})} = (0.007, 1.000, -0.008)$. Similarly, from the MS2 fitting, $\hat{\mu}_0^{(\text{MS2})} = (0.006, 1.000, 0.006)$. The least-squares (LS) fit of Schulz et al. (2015) also provides a similar estimate. These estimates are virtually the same, only 0.6 degrees away from the ground truth μ_0^*. Estimates of the concentric small-circles $\mathcal{C}(\hat{\mu}_0^{(\text{MS2})}, \hat{\nu}_j)$ for four choices of j (the spoke index) are also shown in the top right panel of Fig. 8 in which $\hat{\mu}_0^{(\text{MS2})}$ and μ_0^* are also shown. Although all methods provide virtually the same estimate, only by assuming one of the iMS1, iMS2 or MS2 models, we are able to infer on the axis of rotation. For example, under the iMS1 model, we tested $H_0 : \mu_0 = \mu_0^*$, and with the p-value 0.26, we accept that the true axis of rotation is the hypothesized axis μ_0^*.

2.6.4 Inference on horizontal dependence

An advantage of modeling the s-rep spoke directions by the MS2 distribution is the ability of perceiving and modeling the horizontal dependence among directions. As an exploratory step, we have collected the estimated correlation coefficients, computed from the approximate precision matrix $\widehat{\Sigma}^{-1}$, whose elements are $\hat{\boldsymbol{\kappa}}$ and $\widehat{\boldsymbol{\Lambda}}$; see (2.16). A histogram of $K(K-1)/2$ estimated correlation coefficients is plotted in the bottom left panel of Fig. 8. Notably, pairs of spoke vectors from the same side (e.g. two spoke vectors in the "left side" of the ellipsoids in Fig. 8) exhibit strong positive correlations, while those from the opposite sides exhibit strong negative correlations. The horizontal dependence is in fact apparent by the way data were generated (simultaneously bending all the spoke directions).

We point out that, due to obvious dependence among multivariate directions, to date, the MS2 is the only meaningful model for this data. The iMS1 and the multivariate extension of the BM are not capable of modeling such association. We check that when goodness-of-fit tests based on Jupp (2005) and Székely and Rizzo (2013) are applied to the s-rep data, the tests reject all other distributions except the MS2. See Section 2.8.6 for details.

For large enough sample sizes, we could use the test of association discussed in Section 2.4 for testing $H_0 : \boldsymbol{\Lambda} = \mathbf{0}$. Unfortunately, due to our small sample size, $n = 30$, and the large number of parameters tested, 1653 ($= K(K-1)/2$), this is infeasible. Coping with this high-

dimension, low-sample-size situation is beyond the scope of current paper, and we resort to choosing only two spoke directions to test the dependence, but to repeating the testing for many different pairs of total $K = 58$ spokes. For each pair of spokes, the likelihood-ratio test produces a p-value for the pair. Investigating the empirical distribution of these p-values can provide a rough estimate of the power. In Fig. 8(d), it can be seen that, at the significance level 0.05, the MS2 test of dependence is indeed powerful, with a rejection rate of 98.5%.

To provide some context to this rate, the MS2 test was compared with other natural choices of tests. We applied two methods that were previously used for s-rep data analysis: the composite principal nested spheres (CPNS), discussed in Pizer et al. (2013), and the least-square (concentric) small-circle fitting method of Schulz et al. (2015).

The CPNS-test is built as follows. First, the least-square small-circle is fitted to each marginal direction on $\mathbb{S}^2$. With an understanding that the axis of the fitted small-circle points to the north pole, the observations (say, x_{ik} from the ith sample, kth spoke) are represented in spherical coordinates (θ_{ik}, ϕ_{ik}). For the purpose of testing "horizontal associations", we only keep the longitudinal coordinates θ_{ik}. For any given pair (k, k'), Fisher's z-transformation is used to obtain the p-value in testing whether the correlation coefficient between θ_{ik} and $\theta_{ik'}$ is zero. We refer to this test procedure by a CPNS test.

An LS test procedure is defined similarly to the CPNS test, except that the first step of fitting individual small-circles is replaced by fitting *concentric* small-circles.

These two tests were also conducted for the same combinations of spoke directions, and the empirical distributions of respective p-values are also plotted in Fig. 8. These alternative tests appear to be too conservative, with rejection rates 11% for the LS test, and 11% for the CPNS test (at level 0.05). Heuristically, the higher power of the MS2 test is due to the superior fitting of the MS2 distribution. In particular, the "horizontal angles" predicted from the MS2 tend to be linearly associated, while those from the least-squares fit tend to be arbitrary. We refer to Section 2.8.5 for more numerical results. All in all, using the MS2 distribution shows a clear advantage in modeling and testing the horizontal dependence of multivariate directions.

2.7 HUMAN KNEE GAIT ANALYSIS

In biomechanical gait analysis, accurately modeling human knee motion during normal walking has a potential to differentiate diseased subjects from normal subjects. In particular, the axis of bending (of the lower leg toward the upper leg) is believed to be a key feature in the discrimination among the diseased and normal subjects (Pierrynowski et al., 2010). As a step towards the development of statistical tests for a "two-group axes difference," in this section we employ the proposed distributional families in modeling the bending motion of the knee.

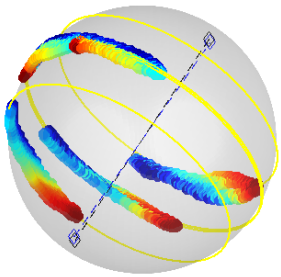
The raw data set we use is obtained from a healthy volunteer and it is a time series of coordinates of markers planted at the volunteer's leg, recorded for 16 gait cycles. For each time point, the directional vectors on $(\mathbb{S}^2)^5$ were computed to be the unit vectors between reference markers, as done in Schulz et al. (2015). As evident from Fig. 9, these directional vectors are *horizontally dependent* of each other, which suggests that we can only fit the MS2 distribution.

The first panel of Fig. 9 illustrates the result of MS2 fit to the all data points ($n = 1000$). There, we superimpose the fitted concentric circles to the observed directional vectors, including their estimated axis, together with a hypothesized dominant bending axis $\mu_0^* = (0, 1, 0)^\top$, the left-right axis of the subject. The axis estimates from the iMS1 or the LS method also provided similar estimates. The MS2 model seems to fit well with high estimated horizontal correlation coefficients. We, however, identify a seemingly strong evidence against using a single MS2. Specifically, as shown in Fig. 9 some directional vectors exhibit higher variations for a subset of time points.

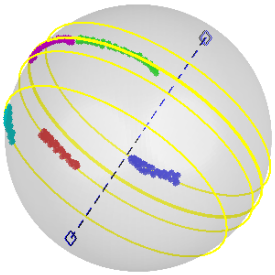
In fact, the data consist of many inhomogeneous periods of the gait cycle. We focus on the "swing" and "stance" periods, and separately analyze subsampled data from each period ($n_{\mathrm{sw}} = 66$ and $n_{\mathrm{st}} = 118$). The MS2 model fits well for the swing period data; see Section 2.8.6 for a goodness-of-fit analysis. The axis of swing is estimated at $\hat{\mu}_0^{(\mathrm{SW})} = (0.013, 1.000, 0.005)^\top$, virtually the same as the hypothesized axis μ_0^*. With the p-value 0.16, we accept that the axis of swing is the left-right axis of this healthy person. For the stance period, excluding the highly-irregular directions shown as dark blue points in Fig. 9(c), the

MS2 model also fits well. We confirm that while in stance, the axis of major rotation differs from μ_0^* with the p-value less than 10^{-5} in testing $H_0 : \mu_0 = \mu_0^*$. The estimated axis for the stance period is $\hat{\mu}_0^{(\mathrm{st})} = (0.11, 0.994, 0.006)^\top$, about 6 degrees away from μ_0^*.

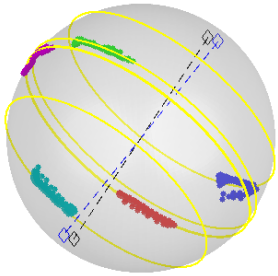
While the MS2 distribution was useful in making inference on the bending axis of partial knee motions, future work for this type of data lies in the development of a two-sample axis difference test.



(a) All data points



(b) Swing period



(c) Stance period

Figure 9: Knee gait data: Observed directional vectors overlaid with the hypothesized (black dashed) and estimated (blue dashed) axes as well as the MS2-fitted small circles. The directions along the north-most circle in the stance period exhibit a higher and irregular pattern of variation. In (a), colors code time indices. In (b) and (c), different colors represent different $\mathbb{S}^2$. The data are horizontally associated, with correlation coefficients ranging from 0.17 to 0.97 (in absolute values), which are significantly different from zero.

2.8 TECHNICAL DETAILS AND SUPPORTING MATERIALS

2.8.1 Proofs of lemmas and propositions

We provide a technical lemma, referenced in Section 2.3.1, and proofs of Propositions 1 and 2.

Lemma 3. *If $X \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$, then $E(X)$ is a linear combination of μ_0 and μ_1.*

Proof of Lemma 3. Suppose that for some $a, b, c \in \mathbb{R}$, $v \in S^{p-1}$ which does not lie in the span of μ_0 and μ_1, $E(X) = a\mu_0 + b\mu_1 + cv$. Then choose a $B \in O(p)$ such that $B\mu_0 = \mu_0$, $B\mu_1 = \mu_1$ but $Bv \neq v$. By Proposition 1(i), $BX \sim S1(\mu_0, \mu_1, \kappa_0, \kappa_1)$. Thus $E(X) = E(BX)$, which in turn leads to $a\mu_0 + b\mu_1 + cv = a\mu_0 + b\mu_1 + cBv$, which is true only if $c = 0$. This gives the result. □

Proof of Proposition 1. Note that if $X \sim \mathrm{S1}(\mu_0, \mu_1, \kappa_0, \kappa_1)$, then $BX \sim \mathrm{S1}(B\mu_0, B\mu_1, \kappa_0, \kappa_1)$. The argument is true when S1 is replaced by S2. Claim (ii) is verified by comparing respective density functions. In (i), verifying the sufficient condition is trivial, thus omitted.

Now suppose that X and BX both have the $\mathrm{S1}(\mu_0, \mu_1, \kappa_0, \kappa_1)$ distribution. Using the parametrization of (2.10), in particular $\gamma = 2\kappa_0\nu\mu_0 + \kappa_1\mu_1$, we have for any $x \in \mathbb{S}^{p-1}$, $(B\gamma)^\top x = \gamma^\top x$ and $\left((B\mu_0)^\top x\right)^2 = (\mu_0^\top x)^2$, which in turn leads to $B\gamma = \gamma$ and $B\mu_0 = \pm\mu_0$. Plugging in $B\mu_0 = -\mu_0$ into the equation $B\gamma = \gamma$, we get

$$\|B\mu_1\|^2 = \|4\nu\frac{\kappa_0}{\kappa_1}\mu_0 + \mu_1\|^2 = 1 + 8\nu^2\kappa_0/\kappa_1(1 + 2\kappa_0/\kappa_1) > 1,$$

which contradicts to the assumption that B is orthogonal. Thus $B\mu_0 = \mu_0$, in which case, $B\mu_1 = \mu_1$ as well.

Next suppose that Y and BY both have the $\mathrm{S2}(\mu_0, \mu_1, \kappa_0, \kappa_1)$ distribution. Without loss of generality, suppose $\kappa_0 = \kappa_1 = 1$. From (2.6), we have for any $x \in \mathbb{S}^{p-1}$,

$$-(\mu_0^\top x - \nu)^2 + \frac{\mu_1^\top x - \nu\mu_0^\top x}{\sqrt{(1-\nu^2)(1-(\mu_0^\top x)^2)}} = -(\mu_0^\top Bx - \nu)^2 + \frac{\mu_1^\top Bx - \nu\mu_0^\top Bx}{\sqrt{(1-\nu^2)(1-(\mu_0^\top Bx)^2)}}. \quad (2.29)$$

Plugging in $x = \pm\mu_0$ in (2.29) yields

$$-(1-\nu)^2 = -(a-\nu)^2 + (b-\nu a)/\sqrt{(1-\nu^2)(1-a^2)},$$
$$-(1+\nu)^2 = -(a+\nu)^2 - (b-\nu a)/\sqrt{(1-\nu^2)(1-a^2)},$$

where $a = \mu_0^\top B\mu_0$, $b = \mu_1^\top B\mu_0$. Solving the above system of equations, we get $a = \mu_0^\top B\mu_0 = \pm 1$, thus $B\mu_0 = \pm\mu_0$. Suppose $B\mu_0 = -\mu_0$, so that (2.29) becomes

$$\left((B\mu_1)^\top - \mu_1^\top\right)x = c\mu_0^\top x, \quad c = \nu^{-1}(\sqrt{(1-\nu^2)(1-(\mu_0^\top x)^2)} - 1),$$

which implies that $B\mu_1 - \mu_1$ is parallel to μ_0. However, the coefficient c is not a constant function of x, and there exists an x such that $\|\mu_0\| = c^{-1}\|B\mu_1 - \mu_1\| \neq 1$, which contradicts the fact that $\mu_0 \in \mathbb{S}^{p-1}$. Thus $B\mu_0$ must be μ_0, in which case we have $B\mu_1 = \mu_1$ as well. □

Proof of Proposition 2. For a given $h > 0$, let $\gamma = 2\kappa_0\nu\mu_0 + \kappa_1\mu_1$ and $A_h = \kappa_0\mu_0\mu_0^\top + hI_p$. Then the S1 density (2.5) can be expressed as the Fisher-Bingham form (2.10):

$$f_{\mathrm{S1}}(x;\mu_0,\mu_1,\kappa_0,\kappa_1) = \alpha(\gamma, A_h)\exp\{\gamma^\top x - x^\top A_h x\}$$

where $\alpha(\gamma, A_h)$ satisfies

$$a(\kappa_0,\kappa_1,\nu) = \alpha(\gamma, A_h)\exp\{-\kappa_0\nu^2 + h\}. \tag{2.30}$$

For the purpose of evaluating the value of $a(\kappa_0,\kappa_1,\nu)$, or equivalently $\alpha(\gamma, A_h)$ for the given value of h, one can assume without losing generality that $\mu_0 = (1,0,\ldots,0)^\top$ and $\mu_1 = (\nu, \sqrt{1-\nu^2}, 0, \ldots, 0)$, so that $\gamma = (\nu(2\kappa_0+\kappa_1), \kappa_1\sqrt{1-\nu^2}, 0, \ldots, 0)^\top$, and the vector of diagonal values of A_h is $\lambda := (2(\kappa_0+h), h, \ldots, h)$. The jth element of ξ, in the statement of the proposition, is then given by $\xi_j := \gamma_j/2\lambda_j$. With these notations, Proposition 1 of Kume and Wood (2005) gives

$$\alpha(\gamma, A_h) = 2\pi^{p/2}|A_h|^{-1/2}g(1)\exp\{\xi^\top A_h\xi\}.$$

Hence, by (2.30), we have (2.17). □

2.8.2 Convergence of algorithm for S1 estimation

We confirmed that both of our algorithms in fitting MS1 and MS2 models converge quickly to local minima of the negative log-likelihood function, in all of the data situations we tested.

As an example, the quick convergence of the proposed algorithm in fitting the S1 density for a toy data set is illustrated in Fig. 10. The left panel of the figure shows the data we used, overlaid with the true and estimated parameters (shown as small circles). The right panel shows the fast decay of the negative log-likelihood, evaluated at each iteration.

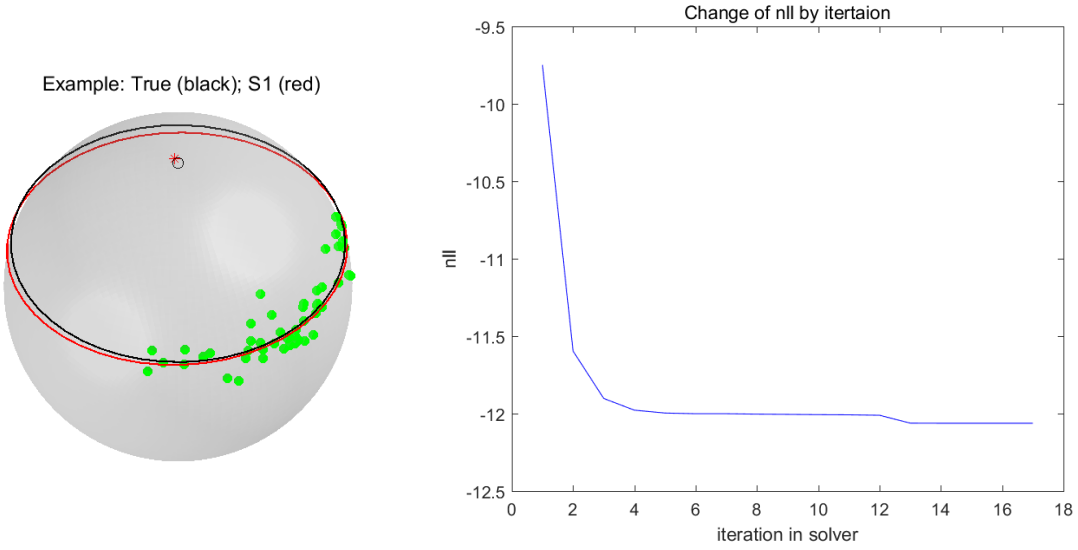


Figure 10: (left) Example data on $\mathbb{S}^2$ following the S1, true parameters (shown as black) and fitted parameters (red). (right) The algorithm of fitting the S1 model requires only 17 iterations to reach the minimum of the negative log-likelihood function.

2.8.3 Null distributions and empirical powers of tests

Null distribution of W_n. As referenced in Section 2.4, we demonstrate that the empirical null distributions of the test statistics W_n are indeed approximately chi-square distributed where we use a moderate sample size $n = 30$. To check this, we use Q-Q envelope plots (Lee, 2007) of test statistics under the null hypotheses. Recall that the parameter space

for the iMS1 and iMS2 models is given by $\Theta_{\text{ind}} = \mathbb{S}^{p-1} \times (\mathbb{S}^{p-1})^K \times (\mathbb{R}_+)^K \times (\mathbb{R}_+)^K$ for $\theta_{\text{ind}} = (\mu_0, \boldsymbol{\mu}_1, \boldsymbol{\kappa}_0, \boldsymbol{\kappa})$. For the more general GMS2 model including associations, we have $\Theta_{\text{GMS2}} = \Theta_{\text{ind}} \times (\mathbb{R}^{(p-1)^2})^{K(K-1)/2}$ for $\theta_{\text{GMS2}} = (\theta_{\text{ind}}, \mathbf{B})$. Note that in Hypothesis 1, using MS2 model, we consider only $p = 3$, whose parameter space is $\Theta = (\theta_{\text{ind}} \times (\mathbb{R})^{K(K-1)/2})$ for $\theta = (\theta_{\text{ind}}, \boldsymbol{\Lambda})$.

1. **Test of association**. H_0: $\boldsymbol{\Lambda} = \mathbf{0}$, or $\theta \in \Theta_0 = \mathbb{S}^{p-1} \times (\mathbb{S}^{p-1})^K \times (\mathbb{R}_+)^K \times (\mathbb{R}_+)^K \times \{\mathbf{0}\}$. Under H_0, the model degenerates to the iMS2 and there is no horizontal dependence.

2. **Test of axis**. H_0: $\mu_0 = \mu_0^*$, or $\theta \in \Theta_0 = \{\mu_0^*\} \times (\mathbb{S}^{p-1})^K \times (\mathbb{R}_+)^K \times (\mathbb{R}_+)^K \times (\mathbb{R})^{K(K-1)/2}$. This is to test whether a predetermined axis μ_0^* of the small sphere is acceptable.

3. **Test of great-sphere**. H_0: $\boldsymbol{\nu} = 0$, or $\theta \in \Theta_0 \simeq \mathbb{S}^{p-1} \times (\mathbb{S}^{p-2})^K \times (\mathbb{R}_+)^K \times (\mathbb{R}_+)^K \times (\mathbb{R})^{K(K-1)/2}$. ($A \simeq B$ means that A and B are diffeomorphic.) This is to test whether the underlying spheres are great spheres with radius 1.

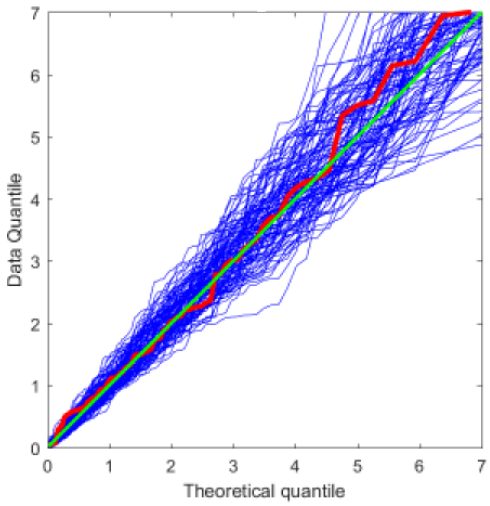
4. **Test for von Mises-Fisher distribution**. H_0: $\kappa_0 = 0$, or $\theta \in \Theta_0 \simeq \mathbb{S}^{p-1} \times \mathbb{R}_+$. Under H_0, there is no "small-circle feature."

5. **Test for Bingham-Mardia distribution**. H_0: $\kappa_1 = 0$, or $\theta \in \Theta_0 \simeq \mathbb{S}^{p-1} \times \mathbb{R}_+$. Under H_0, there is no unique mode.

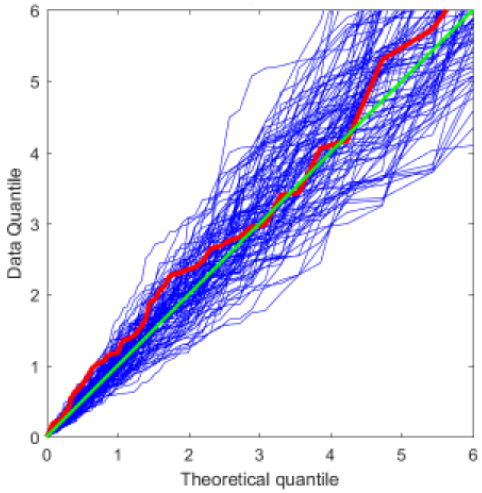
We provide empirical null distributions for Hypotheses 2–5 (labeled as (a)–(d)) while the full model is the S1 distribution. The size of the test of association (Hypothesis 1) can be checked in Fig. 12. The Q-Q envelope plots for each of the test statistics are shown in Fig. 11.

In each panel of Fig. 11, the Quantile-Quantile plot of W_n (with respect to the asymptotic null distribution), simulated under each corresponding null hypothesis, is shown as the red curve. This is overlaid with 100 Q-Q plots (shown in blue curves), obtained from the random samples of the same size, following the theoretical chi-squared distribution. The blue curves provide an envelope, representing the natural variation of χ^2 samples. Based on Fig. 11, arguing as in Lee (2007), we conclude that the test statistic W_n approximately follows the chi-square distribution.

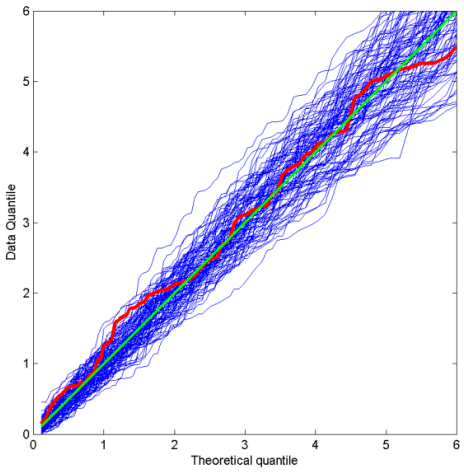
Empirical powers of the tests. We provide empirical powers of the proposed likelihood ratio tests for the test of associations in MS2 (Hypothesis 1) and Hypotheses 3 and 5 under



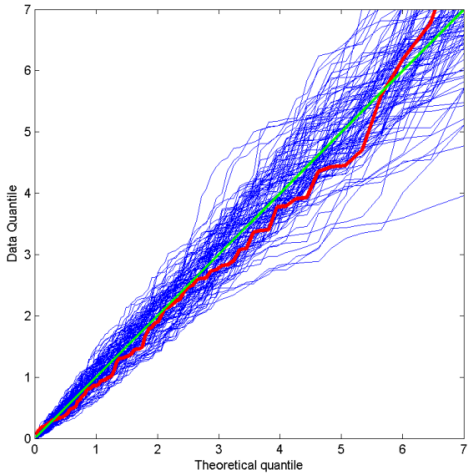
(a) H_0: $\mu_0 = \mu_0^*$



(b) H_0: $\nu = 0$



(c) H_0: $\kappa_0 = 0$



(d) H_0: $\kappa_1 = 0$

Figure 11: Q-Q envelope plots for testing the asymptotic distribution of W_n against χ^2_{df}, where $\text{df} = q_1 - q_2$ denotes the corresponding degrees of freedom. The sample size is $n = 30$. For each case, the Q-Q plot of the test statistics (the red curve) is inside the acceptable variation (given by the envelope within the blue lines). See text and Lee (2007) for the use of the Q-Q envelope plot.

S1. The powers of the test for von Mises-Fisher distribution (Hypothesis 4) are also reported in Section 2.5.3. In all simulations, we used 200 repetitions to compute the empirical rejection rates, at significance level 0.05.

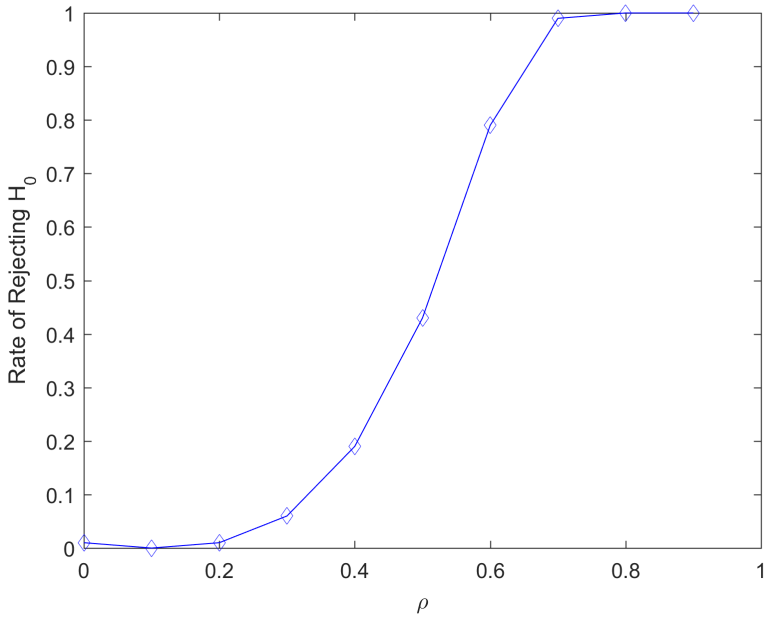
- **Test of association (MS2).** Hypothesis 1. See Fig. 12. This is to test H_0: iMS2 vs. H_1: MS2. In other words, H_0: $\lambda_{jk} = 0$ for all $j \neq k$, $j, k = 1, \ldots, K$. Empirical powers for $K = 2$ and 3 are shown in Fig. 12, for various alternative settings. We reparametrize Λ using the correlation coefficient ρ, as follows. For $K = 2$, the parameters $(\boldsymbol{\kappa}_0, \boldsymbol{\kappa}, \lambda_{12})$ are parameterized by σ_v, σ_h, ρ, representing the vertical standard deviation, horizontal standard deviation, and horizontal correlation coefficient;

$$\kappa_0 = (2\sigma_v^2)^{-1}, \quad \begin{pmatrix} \kappa_1 & -\lambda_{12} \\ -\lambda_{12} & \kappa_2 \end{pmatrix} = \begin{pmatrix} \sigma_h^2 & \rho\sigma_h^2 \\ \rho\sigma_h^2 & \sigma_h^2 \end{pmatrix}^{-1}. \tag{2.31}$$

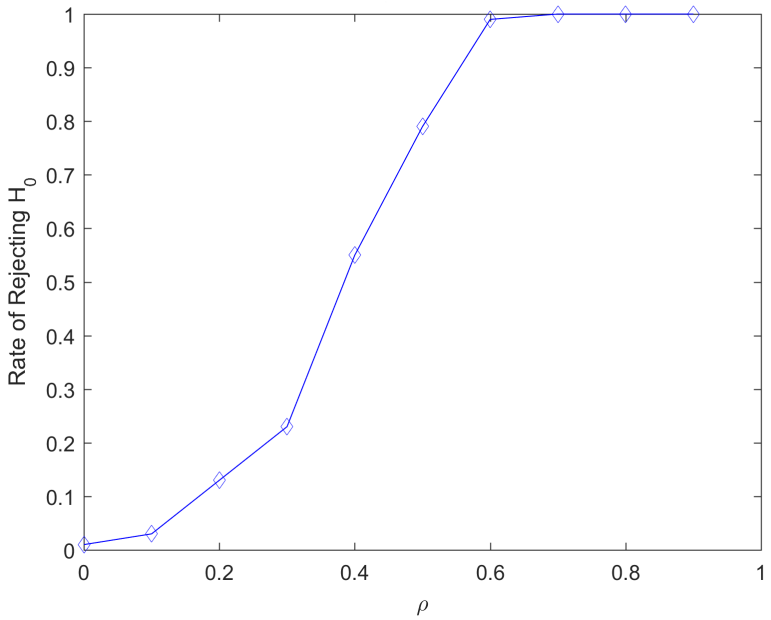
  Figure 12 shows that for the cases $K = 2$ and 3 the power sharply grows from zero to 1, with a notably hike at $\rho = 0.3$. The empirical type I error rates are controlled below the significance level 0.05. The figure is generated for sample size $n = 100$ from the MS2 distribution with the fixed vertical ($\sigma_v^2 = 0.005$) and horizontal ($\sigma_h^2 = 1$) dispersions but with several values of the association parameter ($\rho = 0, 0.1, \ldots, 0.9$).
- **Test of great-sphere (S1).** Hypothesis 3. See Fig. 13. The true parameters are set so that the angles in the small circle are either 90° (i.e., H_0 is true), 80°, 70° and 60° (the latter three fall under the alternative hypothesis H_1). These situations are denoted as cases (a)-(d) in Fig. 13, where we visualized a sample of size 100 for each of the cases, in order to give a visual impression of the "effect size".
- **Test for Bingham-Mardia distribution (S1).** Hypothesis 5. See Fig. 14. In the figure, the null distribution (the BM distribution) is shown in (a). Three different alternative distributions are shown in (b)–(d).

2.8.4 Additional simulation results

As referenced in Section 2.5.1, we provide a supplementary figure and additional simulation results.

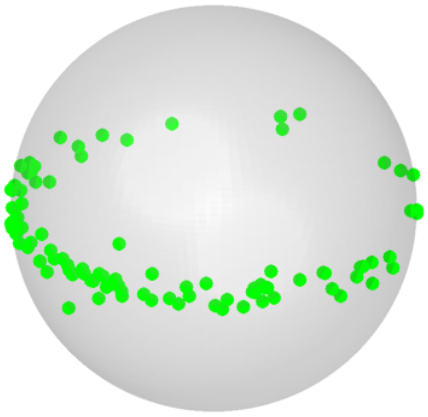


(a) $K = 2$ (MS2)

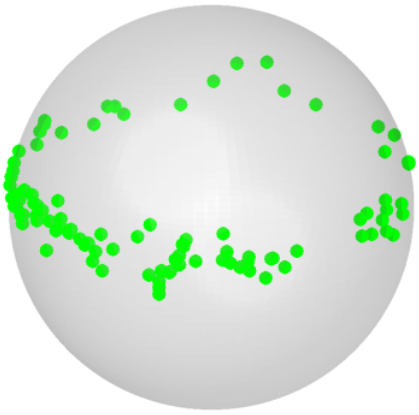


(b) $K = 3$ (MS2)

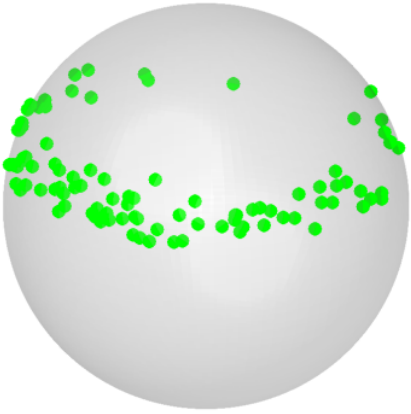
Figure 12: Empirical powers of the test of association at significance level 0.05. The rates of rejecting $H_0 : \lambda_{jk} = 0$ for all $j \neq k$, $j, k = 1, \ldots, K$ computed from 200 repetitions are shown.



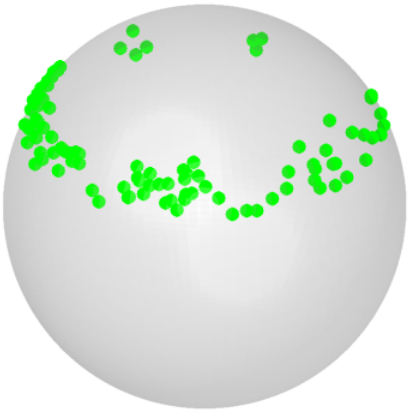
(a) S2 ($\arccos(\nu) = 90°$)



(b) S2 ($\arccos(\nu) = 80°$)

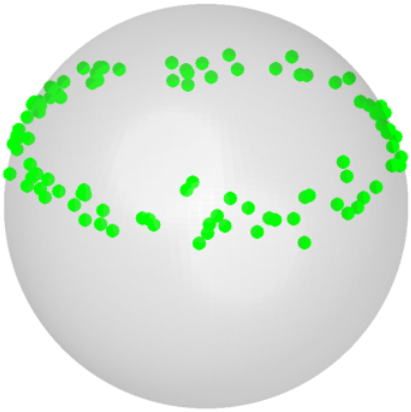


(c) S2 ($\arccos(\nu) = 70°$)

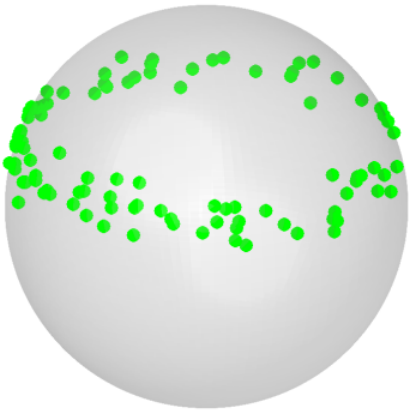


(d) S2 ($\arccos(\nu) = 60°$)

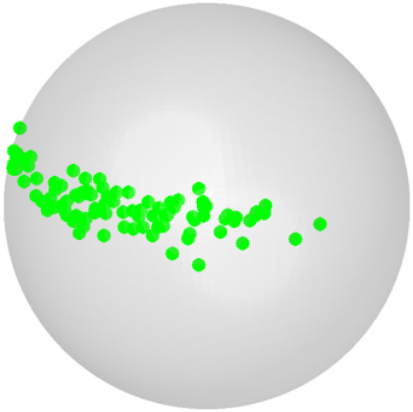
Figure 13: Empirical power $\hat{\beta}$ of testing for a great-sphere (S1): $\hat{\beta} = 0.045$, 1, 1, 1 for cases (a), (b), (c) and (d), respectively. Examples of random samples of size $n = 100$ are shown.



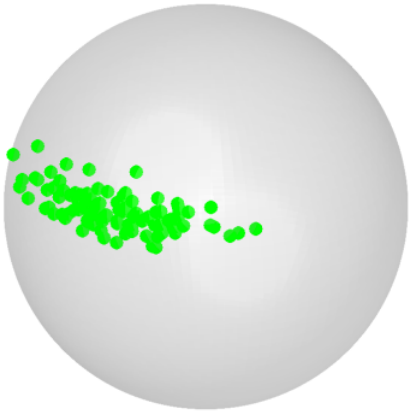
(a) BM ($\kappa = 100$)
S2 ($\kappa_0 = 100, \kappa_1 = 0$)



(b) S2 ($\kappa_0 = 100, \kappa_1 = 0.5$)



(c) S2 ($\kappa_0 = 100, \kappa_1 = 5$)



(d) S2 ($\kappa_0 = 100, \kappa_1 = 10$)

Figure 14: Empirical power $\hat{\beta}$ of testing for the Bingham-Mardia distribution (S1): $\hat{\beta} =$ 0.065, 0.925, 1, 1 for cases (a), (b), (c) and (d), respectively. Examples of random samples of size $n = 100$ are shown.

Bivariate data on $(\mathbb{S}^2)^2$. Figure 15 illustrates random samples from settings (a)–(f) used in Tables 3 and 4.

Robustness against model misspecification. To check robustness against model misspecification, we use a *signal-plus-noise* model that is neither S1 nor S2. This model generates observations on $(\mathbb{S}^2)^K$, that are rotated from a reference point, and perturbed by a spherical noise. Given μ_0, $\mu_j \in \mathbb{S}^2$, the perturbation model for an observation $x_j \in \mathbb{S}^2$ is

$$x_j = R(\mu_0, \theta_j)\mu_j \oplus \epsilon_j \in \mathbb{S}^2, \quad j = 1, \ldots, K, \tag{2.32}$$

where $R(\mu_0, \theta_j)$ is the rotation matrix so that $R(\mu_0, \theta_j)\mu_j$ gives the rotation of the vector μ_j by an angle θ_j about the axis μ_0. The action $\oplus$ is defined as $v \oplus \epsilon = (v + \epsilon)/\|v + \epsilon\|$. The spherical error ϵ_j is independent of $\boldsymbol{\theta}$ and is sampled from $\mathrm{N}(\mathbf{0}, \sigma_v^2 I_3)$ for $\sigma_v^2 > 0$. In the univariate case, θ_1 is sampled from a normal distribution with standard deviation σ_h, and for the bivariate case, (θ_1, θ_2) is sampled from a bivariate normal with the precision matrix specified in (2.31).

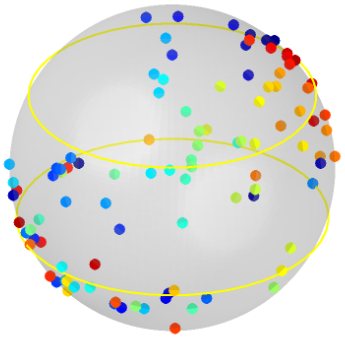
Just like Tables 2 and 3, we report means and standard deviations of the *angular product errors* in Table 5 (univariate cases) and Table 6 (multivariate cases). In Tables 5 and 6, the dispersion parameters $(\sigma_v^2, \sigma_h^2, \rho)$ are carefully chosen so that Case (a) (or b, c, d) of Table 5 corresponds to Case (a) (or b, c, d, respectively) of Table 2. Likewise Case (a) (or b, c, d) of Table 6 corresponds to Case (a) (or b, c, d, respectively) of Table 3.

Our estimates perform rather well against model misspecification. The performance of our estimates is comparable to that of the least-squares estimates, which is designed to estimate the parameters of the signal-plus-noise model. The angular product errors from our estimates (S1, S2, iMS1, iMS2 and MS2) in Tables 5 and 6 are comparable to the least-squares estimates, and are often smaller than the BM estimates.

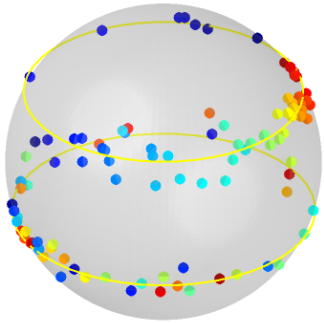
2.8.5 Associations among s-rep spokes

As referenced in Section 2.6.4, we provide more simulation results regarding the test of association, showing the advantage of using the MS2 model in analyzing s-rep data.

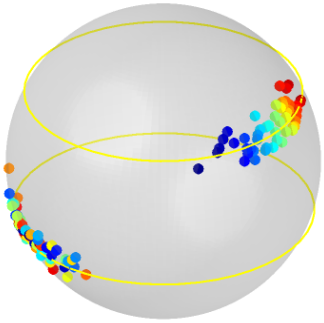
Test of association applied to bi- and tri-variate directions. Previously, we repeatedly applied the test of association, and its competitors, for a pair of spoke directions (i.e.



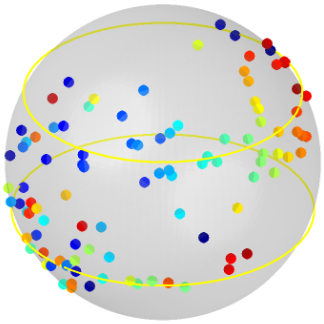
(a) $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (10, 1, 0)$



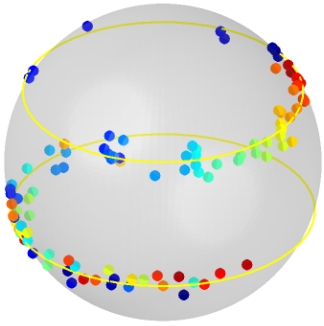
(b) $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (100, 1, 0)$



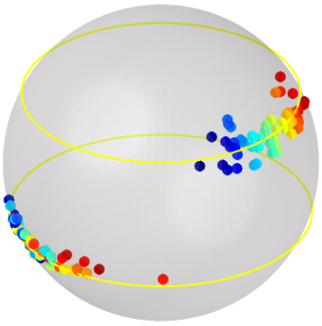
(c) $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (100, 10, 0)$



(d) $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (10, 2, 1.5)$



(e) $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (100, 2, 1.5)$



(f) $(\kappa_{0j}, \kappa_j, \lambda_{12}) = (100, 20, 15)$

Figure 15: Random samples of size $n = 50$ from the MS2 model on $\mathbb{S}^2 \times \mathbb{S}^2$ used in our simulations. Different colors represent different observations. Models (a)–(c) are from iMS2, while models (d)–(f) are from the MS2, with $\lambda_{12} > 0$ (positive association parameters).

Method	(a)	(b)	(c)
S1	7.06(3.74)	1.67(1.16)	10.70(10.77)
S2	7.05(3.89)	1.67(1.08)	**10.32(9.31)**
BM	10.09(7.01)	1.80(1.25)	11.91(10.78)
LS	**6.46(3.50)**	**1.66(1.10)**	10.93(8.96)

Table 5: Small-circle estimation performances for univariate data on $\mathbb{S}^2$. Means (standard deviations) of the angular product errors (in degrees). Data are generated from the signal-plus-noise model (2.32) with $K = 1$.

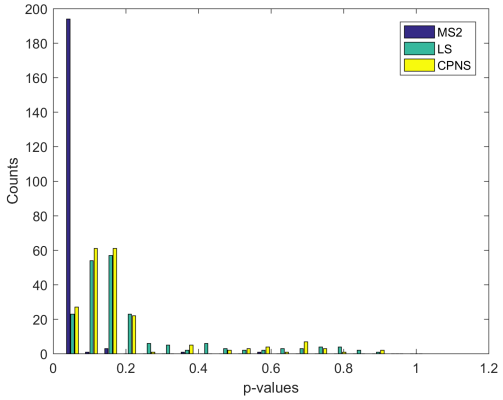
	Independent($\rho = 0$)			Dependent($\rho = 0.7$)		
Method	(a)	(b)	(c)	(a)	(b)	(c)
iMS1	5.22(1.93)	1.51(0.67)	**3.16(3.08)**	5.24(2.16)	1.37(0.59)	**3.49(4.45)**
iMS2	5.15(2.43)	1.50(0.67)	3.78(2.24)	5.03(2.26)	1.36(0.58)	3.62(2.44)
MS2	5.14(2.41)	1.50(0.67)	3.77(2.22)	5.13(2.33)	1.36(0.58)	3.62(2.42)
BM	5.54(2.03)	1.52(0.67)	6.84(3.70)	6.03(3.07)	1.37(0.60)	6.63(3.76)
LS	**4.71(2.05)**	**1.45(0.68)**	3.79(2.26)	**4.73(2.23)**	**1.32(0.58)**	3.65(2.43)

Table 6: Small-circle estimation performances for bivariate data on $\mathbb{S}^2 \times \mathbb{S}^2$. Means (standard deviations) of the angular product errors (in degrees). Data are generated from the signal-plus-noise model (2.32) with $K = 2$.

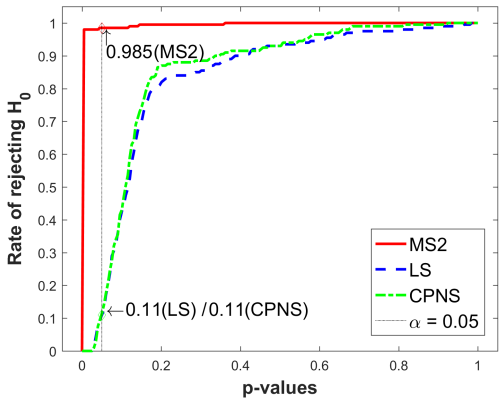
bivariate directions, $K = 2$). We report an additional figure (extending Fig. 8 (d)) for the bivariate case, and results of power comparison for a tri-variate case. For the tri-variate case ($K = 3$), three spokes are randomly chosen among the 58 spokes of the s-rep, and the test of association is applied to compute the p-value for testing $H_0 : \mathbf{\Lambda} = \mathbf{0}$. The two competitors, LS and CPNS tests, discussed in Section 2.6.4, are also applied to obtain corresponding p-values. This is repeated 200 times to obtain an empirical distribution of p-values from the MS2, LS, and CPNS tests of association. Figure 16 summarizes the results. The top row corresponds to the bivariate case (subfigure (b) is also shown in Fig. 8); the bottom row corresponds to the tri-variate case. The left column shows histograms from 200 observed p-values, from the MS2, LS, and CPNS tests of association, respectively. The right column shows the empirical distribution functions (e.d.f.) of the p-values. If the significance level is set at $\alpha \in (0, 1)$, the value of the e.d.f. at α gives the empirical rate of rejection at the significance level α. For example, if $\alpha = 0.05$ in the bivariate case, shown in subfigure (b), the power of the MS2 test of association is estimated at 97%. Notably, the proposed test are much more powerful than the LS test and CPNS test (11 % and 13.5%, respectively). For $K = 3$ case, the power of the MS2 test, 100%, is higher than those of the LS test and CPNS test (10.5 % and 10%, respectively).

Data examples for which MS2 test of association is superior. What makes the MS2 test of association much more powerful than the other two? In Section 2.6.4, we write "*the higher power of the MS2 test is due to the superior fitting of the MS2 distribution.*" To support this claim, we show data examples where the null hypothesis of no association is rejected by the MS2 test, but is not rejected by other tests in Fig. 17 (for the bivariate case) and Fig. 18 (for the tri-variate case). In each of these figures, we show the original data in the left column with different rows corresponding to different small-circle fittings by the MS2 model (top row), the CPNS model (middle row) and the LS model (bottom row). Note that the MS2 and LS models fit concentric circles, while the CPNS model fits circles with generally different axis. In the middle column, the (orthogonal) projections of the data to the corresponding small-circles are shown. In essence, the MS2 test of association and the LS and CPNS tests of correlation are applied to the bi- or tri-variate horizontal angles along the small circles. To give a visual impression of the "linear" association between the

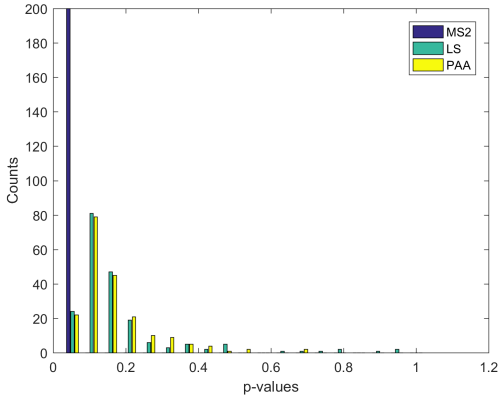
horizontal angles, the scatterplots of the horizontal angles are shown in the right column. In both figures, we check that the small-circle fitting by the MS2 model (top row) provides linearly associated angles. Thus the test of association rejects the null hypothesis of no association. On the other hand, the small-circle fitting by either CPNS (middle row) and LS (bottom row) provides non-linearly associated angles (which is expected, as the both fitting



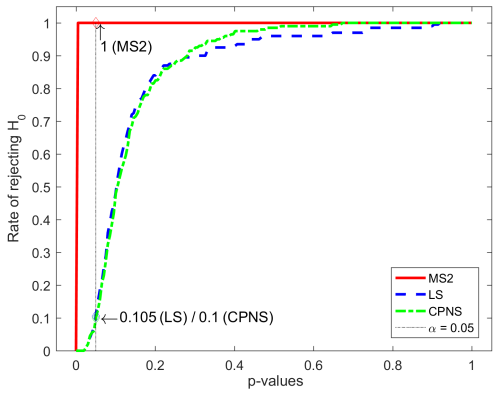
(a) $K = 2$



(b) $K = 2$



(c) $K = 3$



(d) $K = 3$

Figure 16: Histograms ((a) and (c)) and the empirical distribution functions ((b) and (d)) of p-values. The MS2 test exhibits higher powers than LS and CPNS tests.

procedures do not make use of the dependence structure). The test of correlation applied to these data rarely rejects the null hypothesis, resulting in the low powers of the LS and CPNS tests of association.

2.8.6 Goodness of fit

In this section, we briefly report the results of goodness-of-fit analysis for fitting the s-rep data, discussed in Section 2.6, by the MS2 distribution. A bivariate directional data set is used in fitting the parameters of the MS2, iMS2, the BM (our multivatiate extension of the Bingham-Mardia distribution) and the product measure given by two independent von Mises-Fisher distributions.

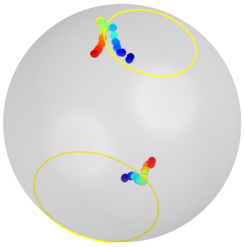
The goodness-of-fit from these distribution families are validated in two ways. First, we visually compared the raw data (on $\mathbb{S}^2 \times \mathbb{S}^2$) with a Monte Carlo sample of the same size ($n = 30$) from each fitted distribution. As the directions are horizontally dependent, only the random sample from the MS2 is visually similar to the raw data; see Fig. 19.

Next, we test the null hypothesis that the raw data are from the given family of distribution. We chose to use the Sobolev test of goodness-of-fit (Jupp, 2005, 2008) that is designed to test the goodness-of-fit for any distribution family on a compact manifold, and a test based on the energy statistic (Székely and Rizzo, 2013).

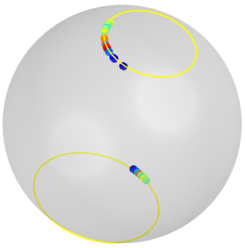
In testing the null hypothesis that the data are from the MS2 distribution, the Sobolev test statistic is computed as follows. Let $\hat{\theta}$ be the MLE of the parameters of the MS2 computed from the data $\{\mathbf{x}_i : i = 1, \ldots, n\}$ where for each i, $\mathbf{x}_i = (x_{i1}, x_{i2})$. The weighted Sobolev statistic is

$$T = \frac{1}{n} \left\| \sum_{i=1}^{n} \frac{\mathbf{t}(\mathbf{x}_i)}{f(\mathbf{x}_i; \hat{\theta})} \right\|^2,$$

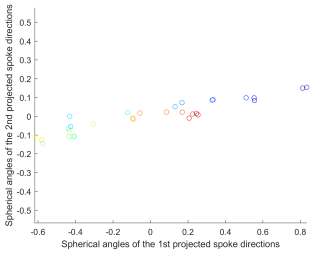
where $\langle \mathbf{t}(\mathbf{x}_i), \mathbf{t}(\mathbf{x}_j) \rangle = P_1(x_{i1}^T y_{i1}) + P_1(x_{i2}^T y_{i2})$, and $P_1(a)$ is the Legendre polynomial of degree 1. For the null distribution of the Sobolev test statistic, we used bootstrapped samples from the fitted distribution, following the suggestion of Jupp (2005). For the data shown in Fig. 19, the test accepts the hypothesis of MS2 distribution with p-value 0.91. While the the hypothesis of BM distribution is rejected with p-value 0, we found that the Sobolev test has no power against the iMS2 distribution, with p-value 0.82. Although we did not test



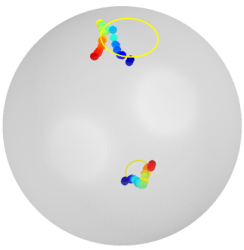
(a) MS2 - Original data and estimates



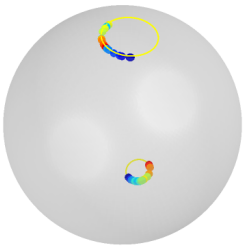
(b) MS2 - Projected data



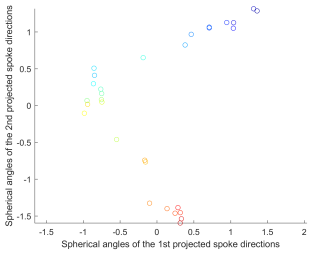
(c) MS2 - Scatter plot of horizontal angles



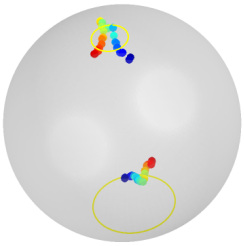
(d) CPNS - Original data and estimates



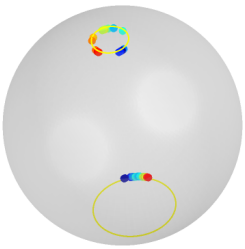
(e) CPNS - Projected data



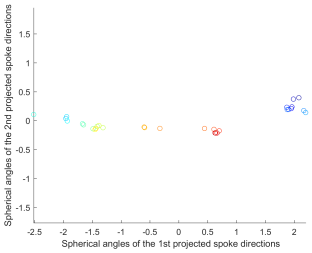
(f) CPNS - Scatter plot of horizontal angles



(g) LS - Original data and estimates

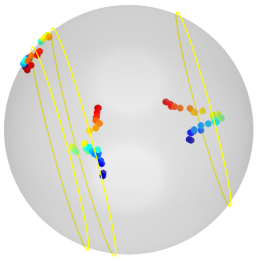


(h) LS - Projected data

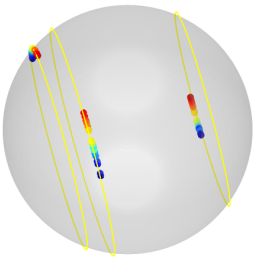


(i) LS - Scatter plot of horizontal angles

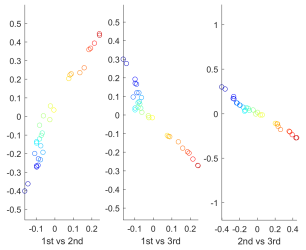
Figure 17: Data example of MS2, CPNS, and LS fittings for a bivariate data set, for which the MS2 test rejects the null hypothesis of no association, while the LS and CPNS tests do not reject. Bright yellow arcs are on the front side of the sphere, and darker yellow arcs are on the back side of the sphere. The horizontal angles predicted from the MS2 tend to be linearly associated, while those from the LS and CPNS fit tend to be arbitrary.



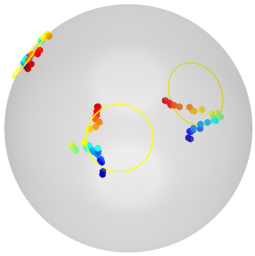
(a) MS2 - Original data and estimates



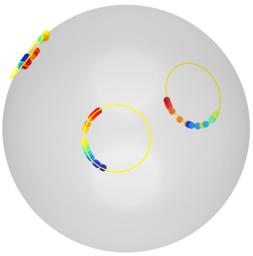
(b) MS2 - Projected data



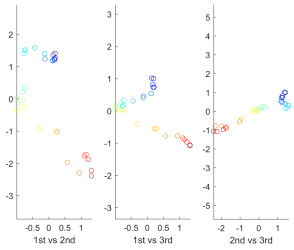
(c) MS2 - Scatter plot of horizontal angles



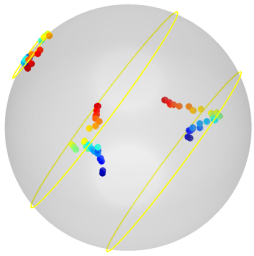
(d) CPNS - Original data and estimates



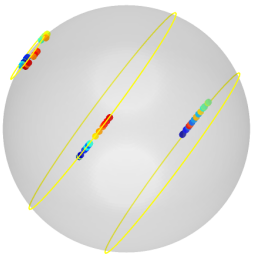
(e) CPNS - Projected data



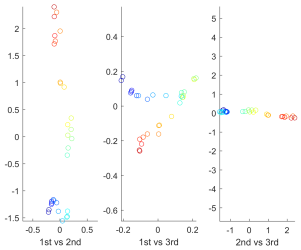
(f) CPNS - Scatter plot of horizontal angles



(g) LS - Original data and estimates



(h) LS - Projected data



(i) LS - Scatter plot of horizontal angles

Figure 18: Data example of MS2, CPNS, and LS fittings for a trivariate data set, for which the MS2 test rejects the null hypothesis of no association, while the LS and CPNS tests do not reject. Bright yellow arcs are on the front side of the sphere, and darker yellow arcs are on the back side of the sphere. The horizontal angles predicted from the MS2 tend to be linearly associated, while those from the LS and CPNS fit tend to be arbitrary.

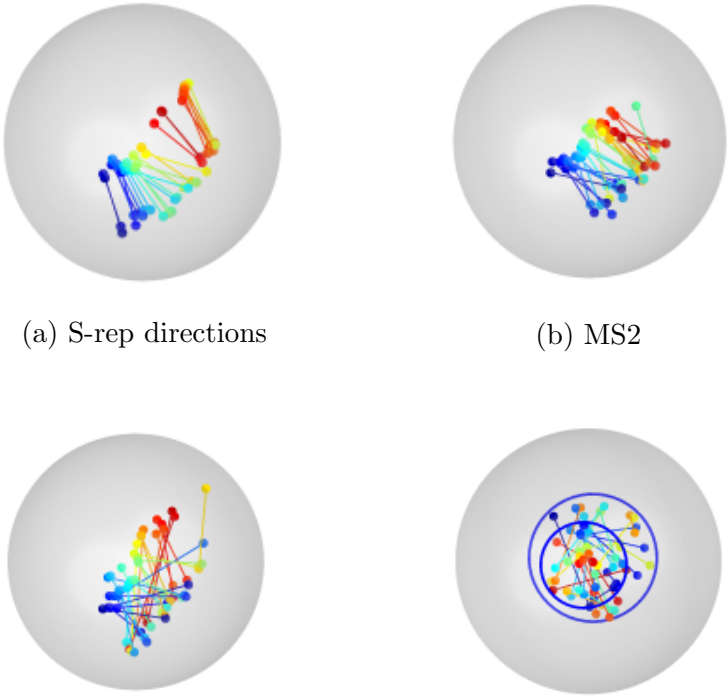
the iMS1 distribution, due to much longer computation time needed in random sampling, we believe the result will be similar to that of the iMS2 distribution.

To further investigate the goodness-of-fit for the iMS2 distribution, we utilized the energy statistic (Székely and Rizzo, 2013), defined for a two-sample comparison. Specifically, in testing the goodness-of-fit for the MS2 distribution, a random sample $\{\mathbf{y}_i\}$ of size $n_* = 60$ is obtained from the MS2 distribution fitted from the data $\{\mathbf{x}_i\}$. The energy statistic is

$$\begin{aligned} e =& \frac{nn_*}{n+n_*}\Big(\frac{2}{nn_*}\sum_{i=1}^{n}\sum_{j=1}^{n_*}\|\mathbf{x}_i-\mathbf{y}_j\| \\ & -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{x}_i-\mathbf{x}_j\| - \frac{1}{n_*^2}\sum_{i=1}^{n_*}\sum_{j=1}^{n_*}\|\mathbf{y}_i-\mathbf{y}_j\|\Big). \end{aligned}$$

The null distribution of e given $\{\mathbf{y}_i\}$ is obtained by permuting the membership to $\{\mathbf{x}_i\}$ or $\{\mathbf{y}_i\}$ of $n+n_*$ observations. Since the corresponding p-value depends on the random sample $\{\mathbf{y}_i\}$, the above procedure is repeated for $R = 100$ independent random samples $\{\mathbf{y}_i\}$, and we use the average of R p-values as the p-value of the test in testing H_0: data follow the MS2 distribution. For the data shown in Fig. 19, the test accepts the hypothesis of MS2 distribution with p-value 0.23. The p-value for the hypothesis of iMS2 distribution turns out to be 0.10, which is at the border between acceptance and rejection with significance level 0.1. The hypotheses of BM and von Mises-Fisher distribtions are rejected with p-value 0.

The analysis is repeated for a number of other pairs of directions. For all pairs we tested, the visual comparison between the raw data and a random sample from the fitted density leads to a similar conclusion that was made by inspecting Fig. 19. The goodness-of-fit tests provided a mixed result. While the hypotheses of BM and von Mises-Fisher distribtions are rejected for all pairs, there are cases where the MS2 distribution is also rejected. However, in each case, the p-value from testing the MS2 is larger than the p-values from testing the BM. The analysis is also repeated for the knee data, discussed in Section 2.7. The results are similar; the p-value for the MS2 is 0.19, accepting the MS2 distribution, and the p-values for the iMS2, BM are respectively 0.06, 0, and 0, showing that the MS2 is the only distribution that fits well to the data.



(a) S-rep directions

(b) MS2

(c) iMS2

(d) BM

Figure 19: (a) A pair of raw s-rep directions on $\mathbb{S}^2 \times \mathbb{S}^2$). Different colors represent different observations. Two directions from each observation are joined by a line segment to emphasize the apparent association among directions. (b) a random sample from the fitted MS2 distribution. (c) a random sample from the fitted iMS2 distribution. (d) a random sample from the fitted BM distribution. Here, we superimposed the BM-fitted small circles. Since it is clear that the data do not follow the von Mises-Fisher (vMF) distribution, we omit the visualization of the vMF sample.

3.0 TEST OF MODALITY FOR ROTATIONALLY SYMMETRIC DISTRIBUTIONS BY MULTIVARIATE KURTOSIS

3.1 INTRODUCTION

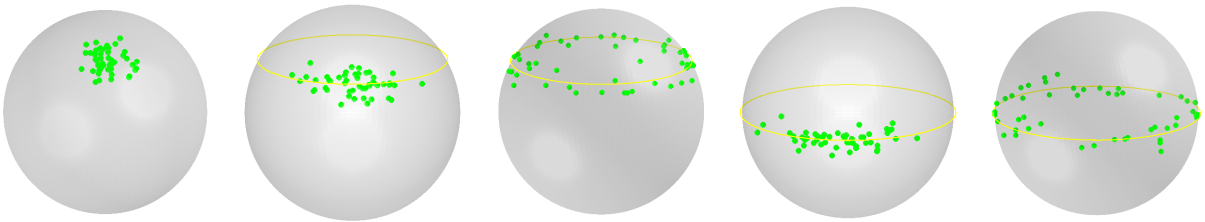
Analysis of principal nested spheres (PNS) proposed by Jung et al. (2012) is a flexible dimension reduction method for directional data, which is an extension of principal component analysis to spheres. In the PNS, the dimension reduction is an iterative procedure of discarding unimportant dimensions. It is specifically designed to capture a certain type of non-geodesic variation by fitting a small sphere.

Figure 20 shows typical data situations on the unit 2-sphere, to which the PNS procedure is applied. These examples are random samples from several spherical distributions; a von Mises-Fisher (vMF; Mardia and Jupp, 2000, p.168) for (a), a small sphere distribution of the second kind (S2) introduced in Chapter 2 for (b) and (d), a Bingham-Mardia (BM; Bingham and Mardia, 1978) for (c) and (e). To visualize the situations, we assumed the dimension as 2, but it can be straightforwardly generalized to higher dimension. The first case indicates that directional vectors are concentrated around a point (center) without any major direction of variation. The second and third cases represent the situation that directional vectors show a non-geodesic major variation while the fourth and fifth cases show a geodesic major variation. The difference between the second and third cases (also, between the fourth and fifth cases) is that directional vectors in the second (and the fourth) case have a single mode while there is no mode in the third (and the fifth) case in which case the distribution is rotationally symmetric.

Since the PNS procedure captures a non-geodesic variation by fitting a small sphere,

the data situations in Figure 20 (b) and (c) are the cases where the PNS is most beneficial compared with other geodesic based dimension reduction methods (Fletcher et al., 2004; Huckemann and Ziezold, 2006; Huckemann et al., 2010; Kenobi et al., 2010). For other cases, a small sphere fit is undesirable because there is no major variation or a geodesic major variation. To treat these cases appropriately, there have been several approaches suggested. First, Jung et al. (2012) adapted a sequence of tests to detect both no major variation and a geodesic major variation cases. As an alternative approach, Jung et al. (2012) also utilized the Bayesian information criterion under Gaussian assumption. Eltzner et al. (2017) proposed another parametric test approach to prevent a small-sphere-fit when a cluster of directional vectors are concentrated around a single center (corresponds to the case in Figure 20 (a)). Because our proposed method in this chapter aims to alternate the existing approaches, we briefly review these and indicate limitations in Section 3.2.

The rest of this chapter is organized as follows. While reviewing existing methods in Section 3.2, cases of which a small sphere fit is treated as an 'overfitting' and corresponding hypotheses for testing approaches will be described. In Section 3.3, an intuitive approach will be discussed, which uses the tangent space mapping to test rotational symmetry of underlying distribution, whose hypothesis corresponds to a part of the second case of overfitting described in Section 3.2. In Section 3.4, we propose a new test, *the test of modality*,



(a) vMF (b) S2 (small) (c) BM (small) (d) S2 (great) (e) BM (great)

Figure 20: Random data examples representing possible situations of interest: There is (a) no major variation, (b and c) a non-geodesic variation, (d and e) a geodesic variation.

which uses the measure of multivariate kurtosis introduced by Mardia (1970). The test is specifically designed for rotationally symmetric distributions. By investigating changes of multivariate kurtosis for rotationally symmetric distributions based on modality, the idea of the proposed test is described. We suggest to use a modified version of the sample kurtosis as a test statistic, which is a consistent estimator of population kurtosis when the underlying distribution is rotationally symmetric. The asymptotic distribution of the modified sample kurtosis will also be investigated. The advantage of the proposed test, when applied to the PNS procedure together with the test of rotational symmetry, is highlighted by numerical examples in Section 3.5. In Section 3.6, as supporting information, an investigation of a uniform distribution defined in a geodesic ball on the unit sphere is given. Last but not least, detailed proofs and calculations are given in Section 3.7.

3.2 BACKGROUNDS

In this section, we briefly review several existing methods aiming at preventing overfitting in PNS.

3.2.1 Sequential tests for preventing overfitting in PNS

Jung et al. (2012) considered two cases where a great sphere provides more appropriate fit to the data.

Case I: Goodness-of-fit test

The first case is when the true major variation is along a great sphere as shown in Figure 20 (d) and (e). Note that when the sample is drawn from a continuous distribution, the PNS procedure fits a small sphere whose geodesic radius is not exactly $\pi/2$ with probability 1. In particular, even when the true geodesic radius is $\pi/2$, the fitted radius is close to but not exactly $\pi/2$.

The likelihood ratio test for geodesic radius. To prevent this type of overfitting, the PNS procedure adapted a likelihood ratio test for the geodesic radius r whose hypotheses

are defined by

$$H_{0a} : r = \pi/2 \text{ vs } H_{1a} : r < \pi/2. \tag{3.1}$$

By assuming that the deviations of the samples from the subsphere are independently distributed as $N(0, \sigma^2)$, the likelihoods for the null and alternative hypotheses are calculated and then the test is conducted by using Wilk's theorem.

Case II: Isotropy test

The second case, which is of interest in this work, is where the data points are concentrated around a point, not on a small sphere, so that there is no major variation along any direction. An example of such a case is illustrated in Figure 20 (a). In this case, the best fit frequently provides a small subsphere with unnecessarily small geodesic radius, which does not give a meaningful decomposition.

The bootstrap von Mises-Fisher test for isotropy. To distinguish such a situation and to choose a great/small subsphere appropriately, the PNS procedure provided the second test, for a distribution function F_X of $X \in \mathbb{S}^d$, whose hypotheses are given by

$$H_{0b} : F_X \text{ is an isotropic distribution with a single mode, vs } H_{1b} : \text{ not } H_{0b}. \tag{3.2}$$

A parametric bootstrap test is proposed by Jung et al. (2012) with an assumption of the von Mises-Fisher distribution (Mardia and Jupp, 2000, p.168). By using the MLE of parameters, the test statistic is designed to be large when F_X is neither isotropic nor having a single mode, then p-value is calculated by simulating bootstrap quantiles.

In the PNS procedure, these two tests are applied to each layer of subsphere fitting as follows:

Step 1 : First, the likelihood ratio test for geodesic radius with hypotheses H_{0a} vs H_{1a} is conducted. If H_{0a} is accepted, then the procedure fits a great sphere with $r = \pi/2$ for the current layer and proceed to the next.

Step 2 : If H_{0a} is rejected, then the procedure runs the bootstrap vMF test with hypotheses, H_{0b} vs H_{1b}. If H_{0b} is accepted, then the procedure uses great spheres for all further subsphere fittings.

Step 3 : If H_{0b} is rejected, then the fitted small sphere is used for the current layer and proceed to the next layer.

We now give an example of how the above steps are applied to each of the data situations in Figure 20. For the case (a), in Step 1, the null hypothesis H_{0a} in (3.1) is frequently rejected because the best fit is given by a small sphere with unnecessarily small radius. In Step 2, we accept the null hypothesis H_{0b} in (3.2), thus fit a great circle. A notable point here is that the procedure uses great spheres for all remaining subsphere fittings without further test if H_{0b} is accepted, in other words, if the distribution is determined to be no major variation. For data in (b) and (c), both H_{0a} and H_{0b} are rejected, so a small circle is fitted. For (d) and (e), the null hypothesis H_{0a} in Step 1 is accepted, so a great circle is fitted.

As described above, by adapting these two tests, the PNS procedure seems to cover all data situations of interest successfully. However, there have been some limitations of this procedure. First of all, because the test in Step 2 uses a bootstrap approach, it requires a large number of repetitions to obtain a stable p-value and in practice it takes a long time. In addition, the vMF distribution assumption made for the second test has been considered to be misspecified. Specifically, in the real data set introduced in Chapter 1 (cf. Figure 2), the directional vectors are often distributed without any major direction of variation similar to the case (a) in Figure 20. However, these directions are distributed close to a uniform distribution rather than the von Mises-Fisher distribution. In such a case, the bootstrap vMF test is found to usually reject the null hypothesis H_{0b} in Step 2. This results in fitting a small circle, that is not desirable.

3.2.2 Bayesian information criterion in PNS

Another approach to detect whether a small sphere is overfitted, provided by Jung et al. (2012), utilized the Bayesian information criterion (BIC) under the Gaussian assumption. Like the likelihood ratio test in the sequential tests above, the deviations of the samples from the subsphere are assumed to be independently distributed as $N(0, \sigma^2)$. Then, by using the MLEs of σ^2 under the small sphere model and the great sphere model, the BICs are calculated and compared.

Through various numerical examples, the BIC approach is found to be effective only for Case I described in Section 3.2.1, i.e., for the case when the true major variation appears along a great sphere as shown in Figure 20 (d) and (e). However, the method is found to be poor in detecting Case II with no major variation.

3.2.3 Folded normal test

As an alternative test approach to prevent overfitting in PNS, Eltzner et al. (2017) proposed a test that used the folded normal distribution to define the distribution of the angular distance s of a data point $X \in \mathbb{S}^d$ from the center of the fitted small sphere $p \in \mathbb{S}^d$: $s = d(p, X)$. The marginal distribution of $s \in [0, \pi]$ is given by

$$f(s; p, \rho, \sigma) = \frac{\sqrt{2\pi}\sigma}{\mathcal{C}(\rho, \sigma)} \sin^{d-1}(s)\mathcal{F}(s; \rho, \sigma),$$

which is obtained by truncating the folded normal density $\mathcal{F}(s; \rho, \sigma)$, defined for $s \in [0, \infty)$,

$$\mathcal{F}(s; \rho, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \left(\exp\left(-\frac{(s - \rho\sigma)^2}{2\sigma^2} \right) + \exp\left(-\frac{(s + \rho\sigma)^2}{2\sigma^2} \right) \right).$$

The shape of the folded normal distribution is characterized by the value of a parameter ρ. For $\rho \to \infty$ the density tends to a usual normal distribution centered at $\rho\sigma$. For $\rho \to 0$ the density becomes a halved normal of doubled height. For $\rho \leq 1$ the mode of distribution stays fixed at the origin, then it moves to the right for $\rho > 1$. Hence, for dimension $d = 2$, the distribution of $X \in \mathbb{S}^2$ forms a modal-ridge along a ring when $\rho > 1$ which is the case where a small sphere fit is desired, while $\rho \leq 1$ yields a uni-modal distribution centered at p which has no proper small spherical structure (no major variation). Therefore, the hypotheses H_{0b} and H_{1b} in (3.2) are reformulated by

$$H_{0b} : \rho = 1 \text{ vs } H_{1b} : \rho > 1.$$

If $\hat{\rho}_{MLE} < 1$, the distribution has its maximum at $s = 0$, i.e., at the center of small sphere fit p, which allows us to accept the null hypothesis (so a great subsphere is fitted) without test. If $\hat{\rho}_{MLE} > 1$, the choice of great or small sphere is determined by the likelihood ratio test by using Wilk's theorem.

Eltzner et al. (2017) have shown by a simulation study that the folded normal test is robust even under the null model , the von Mises-Fisher, of Jung et al. (2012). Furthermore, because the test is based on a likelihood ratio test with an asymptotic chi-square distribution of a test statistic, it is clearly faster than the bootstrap von Mises-Fisher test. However, our simulation study in Section 3.5 shows that this test also does not always work well. Specifically, when a set of concentrated data points around a center forms nearly a uniform distribution, the result of test was unsatisfactory.

3.3 TEST OF ROTATIONAL SYMMETRY BY TANGENT SPACE LIKELIHOOD RATIO TEST

In this section, we discuss a different approach to test whether the underlying distribution of directional vectors on $\mathbb{S}^d$ is rotationally symmetric to alternate the bootstrap vMF test and the folded normal test. Suppose a random direction $X \in \mathbb{S}^d$ has a distribution function F_X and a unique (geodesic) mean direction μ. Then, a rotational symmetry, a part of the hypotheses H_{0b} and H_{1b} in (3.2) is tested by follows.

Assume that we now want to perform a test with hypotheses,

$$H_{0b'} : F_X \text{ is rotationally symmetric, vs } H_{1b'} : \text{ not } H_{0b'}. \tag{3.3}$$

To test these hypotheses, an efficient way is to consider the inverse exponential map of X onto the tangent space at the mean direction, $T_\mu \in \mathbb{R}^d$. To facilitate understanding of the approach, we temporarily assume $d = 2$. Now, let μ be the north pole $(0, 0, 1)^\top$ without loss of generality and φ be the exponential map $Exp_\mu : \mathbb{R}^2 \to \mathbb{S}^2$. Then the inverse exponential map φ^{-1} is defined for $x = (x_1, x_2, x_3)^\top \in \mathbb{S}^2$ by

$$\varphi^{-1}(x) = Log_\mu(x) = \left(x_1 \frac{\theta}{\sin\theta}, x_2 \frac{\theta}{\sin\theta} \right),$$

where $\theta = \arccos(x_3)$ is the geodesic distance from the north pole to the point x. When F_X is rotationally symmetric with respect to the (geodesic) mean direction μ, then the image of the inverse exponential map of X, $\varphi^{-1}(X)$, has the same marginal variance for any arbitrary

choice of coordinates for the tangent space. This allows us to re-write the hypotheses (3.3) as

$$H_{0b'} : \Sigma = \sigma^2 I_d \text{ , vs } H_{1b'} : \text{ not } H_{0b'},$$

where Σ is the variance-covariance matrix of $Y = \varphi^{-1}(X)$ and I_d is the d-dimensional identity matrix.

Now, if we assume a normal distribution for Y, then we can use the likelihood ratio test. Let $Y_1, \ldots, Y_n$ ($\in \mathbb{R}^d$) be the images of inverse exponential map of random directions $X_1, \ldots, X_n$ ($\in \mathbb{S}^d$), and assume Y is normally distributed. Then, the log-likelihoods for the null and alternative hypotheses, $\mathcal{L}_0$ and $\mathcal{L}_1$ respectively, are given by

$$\mathcal{L}_0 = -\frac{n}{2}(\log(|s^2 I_d|) + \sum_{i=1}^{n}(y_i - \bar{y})^\top (y_i - \bar{y})/s^2 + dn\log(2\pi)$$
$$\mathcal{L}_1 = -\frac{n}{2}(\log(|S|) + \sum_{i=1}^{n}(y_i - \bar{y})^\top S^{-1}(y_i - \bar{y}) + dn\log(2\pi),$$

where $s^2 = \frac{1}{d(n-1)}\sum_{i=1}^{n}(Y_i - \bar{Y})^\top (Y_i - \bar{Y})$, and S is the sample variance-covariance matrix of Y_i's. The likelihood ratio test statistic is $W_n = -2(\mathcal{L}_0 - \mathcal{L}_1)$, and, by Wilk's theorem, W_n is asymptotically chi-square distributed with degrees of freedom $d^2 - 1$, which is the difference of the numbers of parameters to be estimated. Hence, when $W_n > \chi^2_{\alpha, d^2-1}$, we reject the rotational symmetry.

Because this approach simply tests that the variance-covariance matrix is a scaled identity, it is geometrically intuitive and efficient in computing. In addition, it is not required to assume a specific spherical distribution for the test. However, this test cannot stand alone to distinguish all the cases concerned in Figure 20. Note that it only tests a rotational symmetry, which means that the test effectively distinguishes the case (a) of no major variation from the cases (b) and (d) which have non-geodesic or geodesic major variation with a single mode. But the cases (a), (c) and (e) cannot be separated by this test of rotational symmetry.

3.4 TEST OF MODALITY FOR ROTATIONALLY SYMMETRIC DISTRIBUTIONS

Recall that among the cases in Figure 20, the likelihood ratio test for geodesic radius given in Case I in Section 3.2.1 is used to effectively separate (b, c) and (d, e). In addition, the cases (b) and (d) are successfully separated from others by the test of rotational symmetry given in Section 3.3. So, in this section, we propose a test to discriminate (a) from (c) and (e).

3.4.1 The measure of multivariate kurtosis

Mardia (1970) proposed a measure of multivariate kurtosis and investigated its properties. Let $X_1, \ldots, X_n$ be a random sample from a d-variate population with random vector X having its density f, mean vector μ and covariance matrix Σ. Then the population measure of multivariate kurtosis and the corresponding sample kurtosis are defined by

$$\kappa(f) = E\{(X-\mu)^\top \Sigma^{-1}(X-\mu)\}^2 \tag{3.4}$$

$$k(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n}\{(X_i-\bar{X})^\top S^{-1}(X_i-\bar{X})\}^2, \tag{3.5}$$

where $\bar{X}$ and S are the sample mean vector and covariance matrix, respectively.

The measure of multivariate kurtosis is known to be invariant under the nonsingular transformation of X: $X \mapsto AX+b$ for $A \in GL(d)$ and $b \in \mathbb{R}^d$. By letting $Z = \Sigma^{-1/2}(X-\mu)$, the population measure of multivariate kurtosis (3.4) is re-written by

$$\kappa(f_Z) = E\{Z^\top Z\}^2 = E\{\|Z\|^4\},$$

where $\|\cdot\|$ is the usual L_2-norm. Hence, the measure of multivariate kurtosis can be understood as the fourth moment of the distance of the standardized data points from the origin. This indicates that $\kappa(f)$ is the straightforward extension of kurtosis in the univariate case to a general dimension $d > 2$. Note that, for a univariate random variable Y with its mean μ_Y and variance σ_Y^2, the kurtosis is defined by the fourth standardized moment $E[(Y-\mu_Y)/\sigma_Y]^4$. The sample measure of multivariate kurtosis (3.5) is also re-written by the same way with replacing μ and Σ by $\bar{X}$ and S, provided that S is of full rank.

3.4.2 The multivariate kurtosis for rotationally symmetric distributions

From now on, we discuss the multivariate kurtosis of several distributions of interest. When we consider rotationally symmetric distributions, they can be separated into three cases based on modality as follows:

Modality 1 : A single mode at the center of the distribution, i.e., the distribution is peaked at the center, for example, a normal distribution with $\Sigma = \sigma^2 I_d$.

Modality 2 : No mode (or infinitely many modes), for example, a ball uniform distribution (3.6).

Modality 3 : Modal-ridge along a circle (or sphere), i.e., the distribution is evenly peaked along a circle (or sphere), for example, a Bingham-Mardia distribution (Bingham and Mardia, 1978).

It is intuitively clear, by the definition, that there is no rotationally symmetric distributions having two or more (countable) number of modes. In this section, we investigate the multivariate kurtosis of these distributions representing the three different cases of modality. For simplicity, we will consider the distributions on $\mathbb{R}^d$, rather than on $\mathbb{S}^{d-1}$. These distributions can be considered as the images in a tangent space of rotationally symmetric distributions on the unit sphere $\mathbb{S}^d$, or as good approximations of them. Hence, by investigating the multivariate kurtosis of these distributions, we reach to the idea of testing modality for rotationally symmetric distributions.

The multivariate normal distribution. Let X be a d-dimensional multivariate normal random variable with mean $\mu = \mathbf{0}$ and covariance matrix $\Sigma = \sigma^2 I_d$. By the invariance property of multivariate kurtosis under the nonsingular transformation, the kurtosis does not depend on the mean and covariance matrix. However, since we are considering rotationally symmetric distributions in this section, we assume the form of covariance matrix to be a scaled identity matrix. This multivariate normal distribution represents Modality 1 with a single mode. Mardia (1970) showed that the population multivariate kurtosis of a multivariate normal distribution is given by

$$\kappa(\phi) = d(d+2),$$

where ϕ denotes the pdf of multivariate normal. Note that the kurtosis only depends on the dimension, but not on the parameters μ and Σ of the normal distribution.

The ball and ring uniform distributions. Now, we consider two other distributions which represent Modalities 2 and 3, respectively. First, let $Y = (Y_1, \ldots, Y_d)^\top$ be a d-variate random variable which is uniformly distributed in a d-dimensional ball centered at the origin with a radius $\theta > 0$ whose density is defined by

$$f_\theta(y_1, \ldots, y_d) = \frac{1}{V_\theta} I_{[\sum_{i=1}^d y_i^2 \le \theta^2]}. \tag{3.6}$$

Here, I is the indicator function and V_θ is the volume of d-dimensional ball, $V_\theta = \frac{\pi^{d/2}\theta^d}{\Gamma(d/2+1)}$. Clearly, the distribution is rotationally symmetric and has no mode (or infinitely many modes because all points in the range of the distribution are modes). From now on, we simply call the distribution with its density (3.6) by a *ball uniform* distribution. When $d = 2$, it is the uniform distribution on the disk with radius θ, which is a good approximation for a (geodesic) ball uniform distribution on $\mathbb{S}^2$. This approximation will be discussed in detail in Section 3.6.

Next, let $U = (U_1, \ldots, U_d)^\top$ be a d-variate random variable which is uniformly distributed in a 'hollowed' d-dimensional ball whose density is defined with two radii (θ_1, θ_2) by

$$f_{\theta_1,\theta_2}(u_1, \ldots, u_d) = \frac{1}{V_{\theta_1} - V_{\theta_2}} I_{[\theta_1^2 \le \sum_{i=1}^d u_i^2 \le \theta_2^2]}. \tag{3.7}$$

Here, V_{θ_i} is the volume of ball with the radius θ_i, $i = 1, 2$. When the dimension is 2, as depicted in Figure 21, the distribution is defined in a ring. So, we simply call the distribution with its density (3.7) by a *ring uniform* distribution. We notice that the distribution is rotationally symmetric but has no mode.

The examples in Figure 20 (c) and (e), which represent Modality 3 with a modal-ridge along a circle, are generated from the Bingham-Mardia distribution. We use the ring uniform distribution as a simple approximation of rotationally symmetric distributions with a modal-ridge to investigate how the modality affects the multivariate kurtosis of the distributions. In particular, because the ball uniform distribution is a special case of the ring uniform distribution when $\theta_1 = 0$, it is now possible to explain the change of kurtosis by a few parameters.

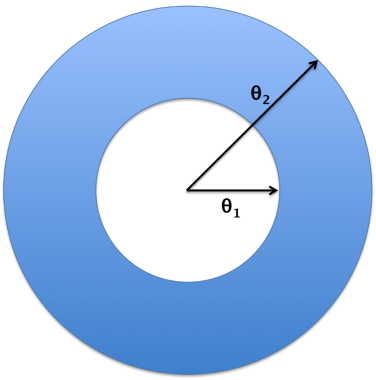


Figure 21: The ring uniform distribution is defined in a hollowed d-dimensional ball, which is a ring when the dimension is 2. As θ_1 getting smaller to 0, the ring uniform distribution is getting closer to the ball uniform distribution.

Proposition 4. *The population measure of multivariate kurtosis for*

1. *the ball uniform distribution is*

$$\kappa(f_\theta) = \frac{d(d+2)^2}{d+4}; \tag{3.8}$$

2. *the ring uniform distribution is*

$$\kappa(f_{\theta_1,\theta_2}) = \frac{d(d+2)^2}{(d+4)} \cdot \frac{(1-\eta^d)(1-\eta^{d+4})}{(1-\eta^{d+2})^2}, \tag{3.9}$$

where $\eta = \theta_1/\theta_2$ $(\in [0,1])$.

Proof. The population multivariate kurtosis (3.4) for the ball and ring uniform distributions can be calculated using the following lemma, a proof of which is given in Section 3.7.1.

Lemma 5. *Let* $U = (U_1, \ldots, U_d)^\top$ *be a* d*-variate ring uniform random variable with radii parameters* (θ_1, θ_2)*. Then,*

1. *for any $j = 1, \ldots, d$ and a given positive integer s, the expected value of U_j^s is given by*

$$E(U_j^s) = \begin{cases} 0 & , \ s \text{ is odd}, \\ d\frac{(\theta_2^{d+s} - \theta_1^{d+s})}{(\theta_2^d - \theta_1^d)} \frac{(s-1)!!}{(d+s)(d+s-2)\cdots(d)} & , \ s \text{ is even}. \end{cases}$$

Here, $c!!$ is the double factorial defined by $c(c-2)\cdots 1$.

2. *For any $j = 1, \ldots, d$ and nonnegative integers s_j's, the expected value of $U_j \prod_{j' \neq j} U_{j'}^{s_{j'}}$ is*

$$E(U_j \prod_{j' \neq j} U_{j'}^{s_{j'}}) = 0.$$

3. *For any $j \neq j'$, $j, j' = 1, \ldots, d$, the expected value of $U_j^2 U_{j'}^2$ is*

$$E(U_j^2 U_{j'}^2) = \frac{(\theta_2^{d+4} - \theta_1^{d+4})}{(\theta_2^d - \theta_1^d)} \frac{1}{(d+4)(d+2)}.$$

Lemma 5 also applies for the ball uniform distribution. All required expectations for the ball uniform distribution can be obtained by setting $\theta_1 = 0$.

Now, the population multivariate kurtosis (3.4) for U is calculated by

$$\begin{aligned}
\kappa(f_{\theta_1,\theta_2}) &= E\{(U - E(U))^\top \Sigma_U^{-1} (U - E(U))\}^2 \\
&= E\{U^\top \Sigma_U^{-1} U\}^2 \\
&= E\{U^\top U\}^2 \left(\frac{(d+2)(\theta_2^d - \theta_1^d)}{(\theta_2^{d+2} - \theta_1^{d+2})} \right)^2 \\
&= E\{\sum_{j=1}^d U_j^4 + \sum_{j \neq j'} U_j^2 U_{j'}^2\} \left(\frac{(d+2)(\theta_2^d - \theta_1^d)}{(\theta_2^{d+2} - \theta_1^{d+2})} \right)^2 \\
&= \left(dE(U_j^4) + d(d-1)E(U_j^2 U_{j'}^2) \right) \left(\frac{(d+2)(\theta_2^d - \theta_1^d)}{(\theta_2^{d+2} - \theta_1^{d+2})} \right)^2 \\
&= \left(\frac{3d(\theta_2^{d+4} - \theta_1^{d+4})}{(d+4)(d+2)(\theta_2^d - \theta_1^d)} + \frac{d(d-1)(\theta_2^{d+4} - \theta_1^{d+4})}{(d+4)(d+2)(\theta_2^d - \theta_1^d)} \right) \left(\frac{(d+2)(\theta_2^d - \theta_1^d)}{(\theta_2^{d+2} - \theta_1^{d+2})} \right)^2 \\
&= \left(\frac{d(\theta_2^{d+4} - \theta_1^{d+4})}{(d+4)(\theta_2^d - \theta_1^d)} \right) \left(\frac{(d+2)(\theta_2^d - \theta_1^d)}{(\theta_2^{d+2} - \theta_1^{d+2})} \right)^2.
\end{aligned}$$

Then, by letting $\eta = \theta_1/\theta_2$, (3.9) follows. Again, we obtain (3.8) for the ball uniform distribution by setting $\theta_1 = 0$ in above calculations. □

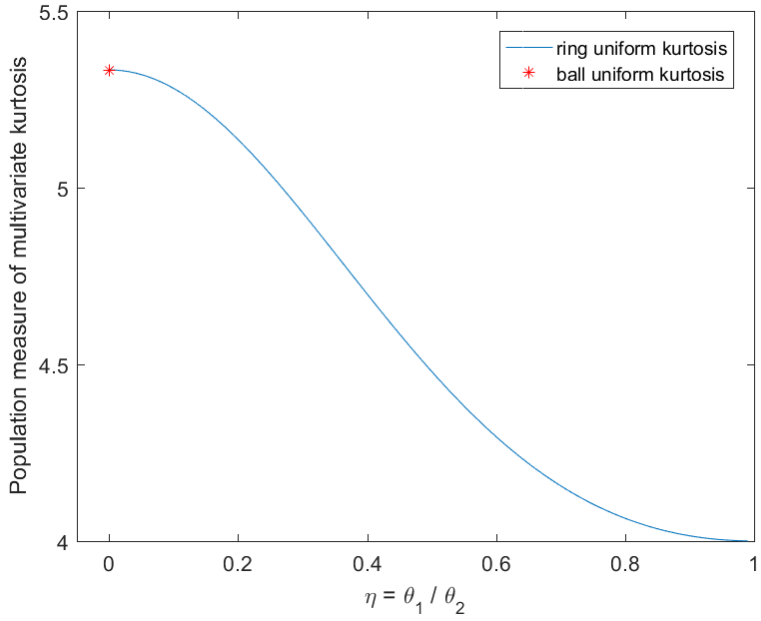


Figure 22: The population multivariate kurtosis for the ball and ring uniform distributions when $d = 2$. The kurtosis for the ring uniform distribution depends on the ratio of the two parameters $\eta = \theta_1/\theta_2$ and increases monotonically as η getting smaller, which eventually converges to the kurtosis for the ball uniform distribution.

We can see that the kurtosis for the ball uniform distribution only depends on the dimension of the distribution, which is intuitively reasonable by the invariance property of the kurtosis. On the other hand, the kurtosis for the ring uniform distribution depends on the ratio of two radii parameters as well as the dimension, which is also intuitive because the covariance matrix Σ depends on these two parameters.

Figure 22 shows how the changes of multivariate kurtosis when the parameter η varies for a fixed dimension $d = 2$. We see that the kurtosis for the ring uniform (3.9) is always smaller than but converges to the kurtosis for the ball uniform (3.8) as $\eta \to 0$.

The convolution and mixture of ball uniform and multivariate normal distributions. To investigate the variation of kurtosis from Modality 1 (a single mode) to Modality 2 (no mode), here we consider two approaches: a convolution of two representative

distributions, a normal and a ball uniform, and a mixture of those. Let $Z = (1-c)X + cY$ for independent X and Y, where X is a d-variate normal random vector with mean 0 and covariance matrix $\Sigma = \sigma^2 I_d$ and Y is a ball uniform random vector with radius θ. For a convolution form, we assume c is a fixed parameter in $[0,1]$. For a mixture, we assume c is a binary random variable with probability $P(c=1) = p$, where $p \in [0,1]$ is given.

Proposition 6. *1. The measure of multivariate kurtosis for a convolution form Z with a fixed parameter $c \in [0,1]$ is given by*

$$\kappa(f_{d,\theta,\sigma,c}) = \frac{\frac{d}{d+4}c^4\tau^4 + d(d+2)(1-c)^4 + 2dc^2(1-c)^2\tau^2}{\left(\frac{1}{d+2}c^2\tau^2 + (1-c)^2\right)^2}, \tag{3.10}$$

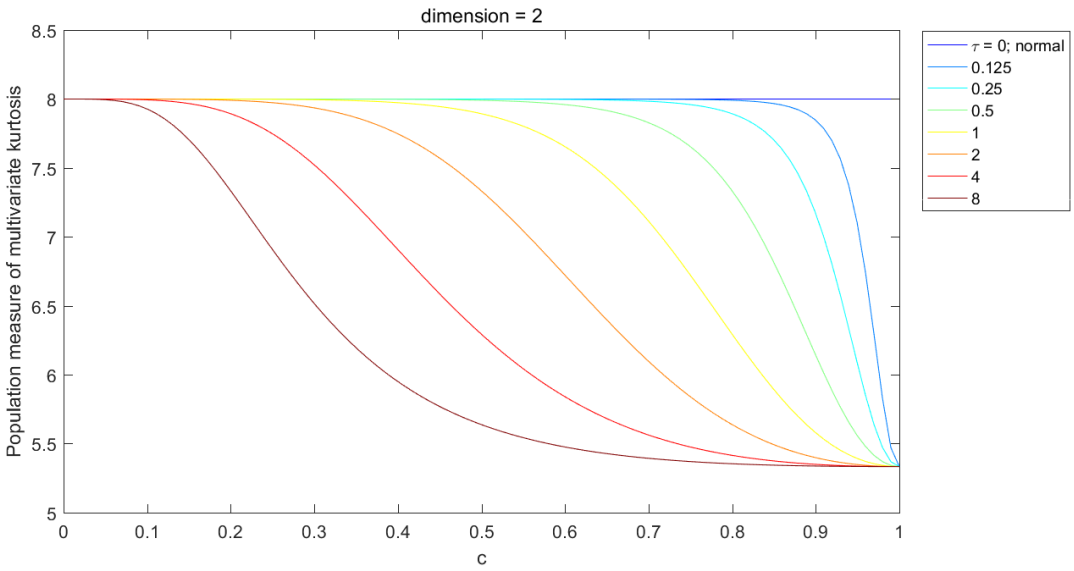
where $f_{d,\theta,\sigma,c}$ denotes the density function of Z and $\tau = \theta/\sigma$ is the ratio of the radius θ of the ball uniform and the standard deviation σ of the normal distribution;

2. The multivariate kurtosis for a normal-ball uniform mixture Z with $P(c=1) = p$ is given by

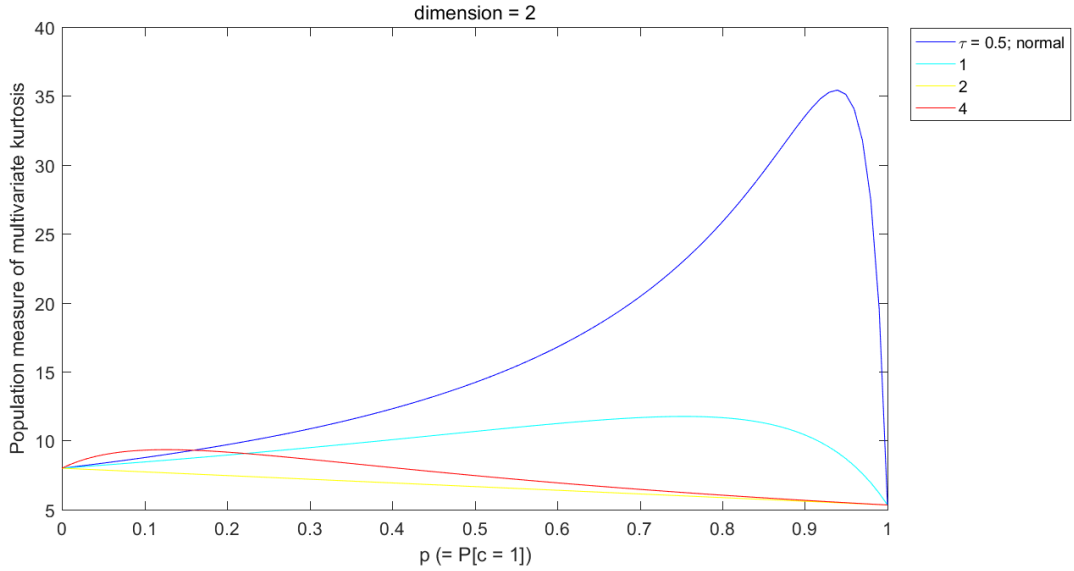
$$\kappa(g_{d,\theta,\sigma,p}) = \frac{\frac{pd\tau^4}{d+4} + (1-p)d(d+2)}{\left(\frac{p\tau^2}{d+2} + (1-p)\right)^2}, \tag{3.11}$$

where $g_{d,\theta,\sigma,p}$ denotes the density function of Z.

The proof of Proposition 6 is given in Section 3.7.2. Figure 23 (a) and (b) show changes of kurtosis (3.10) and (3.11), respectively, for $d = 2$ by changing the the given constant c or the given probability p. As a result, we confirm that the measure of multivariate kurtosis for the ball uniform population is always smaller than any other rotational symmetric and uni-modal distributions. The change of population multivariate kurtosis for the cases of rotationally symmetric distributions based on modality, in particular, from Modality 1 to Modality 3, is summarized in Figure 24. This relationship allows us to consider the kurtosis as a test statistic to test whether the underlying rotational symmetric distribution is uni-modal or not.



(a) Convolution



(b) Mixture

Figure 23: The population measure of multivariate kurtosis for a convolution and mixture of normal and ball uniform distributions when $d = 2$. The distribution of Z is a normal when $c = 0$ in (a) and $p = 0$ in (b), and a ball uniform when $c = 1$ in (a) and $p = 1$ in (b). The kurtosis for the ball uniform distribution is always less than any other convolution or mixture with different τ, c and p.

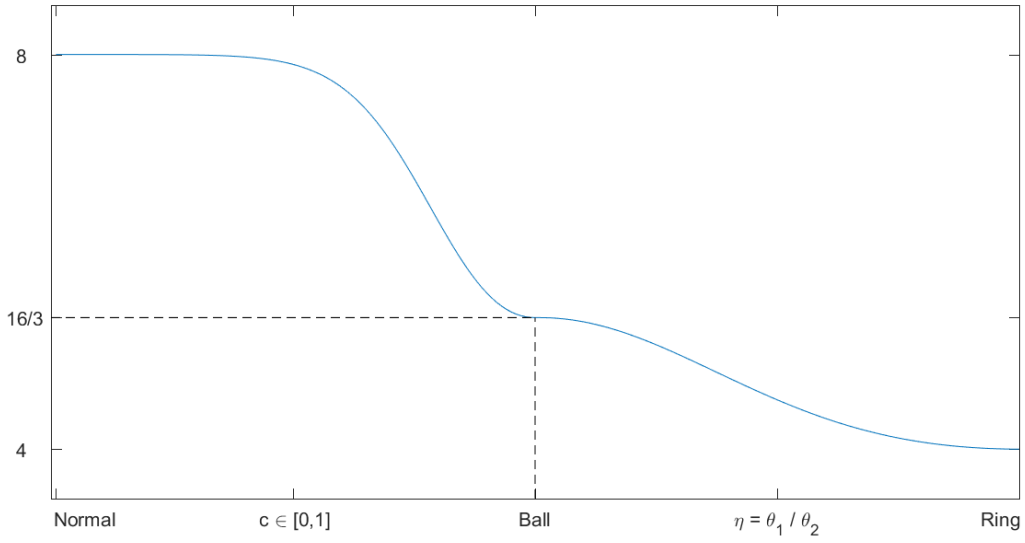


Figure 24: The change of population multivariate kurtosis from the first case (a single mode at center; normal) to the third case (modal-ridge along a circle; ring uniform) when $d = 2$.

3.4.3 Test statistic and its asymptotic sampling distribution

So far, we investigated the variation of population multivariate kurtosis for rotationally symmetric distributions representing the cases of interest. In this section, we propose a modified version of sample kurtosis, which is specifically considered for rotationally symmetric distributions, as a statistic for the test of modality, and investigate its asymptotic sampling distribution.

3.4.3.1 Modified sample kurtosis for rotationally symmetric distributions

Recall that, in the test of rotational symmetry in Section 3.3, we tested whether the covariance matrix of the underlying distribution is a scaled identity or not. In particular, we used a multivariate normal distribution with its covariance matrix $\Sigma = \sigma^2 I_d$ to represent rotationally symmetric distributions with a single mode at the center (Modality 1). Furthermore, by Lemma 5, we can show that the ball and ring uniform distribution also have their covariance

matrices with scaled identities, $\frac{\theta^2}{d+2} I_d$ and $\frac{\theta_1^2+\theta_2^2}{d+2} I_d$, respectively.

This allows us to modify the sample kurtosis by using $\tilde{S} = s^2 I_d$ with the pooled sample variance $s^2 = \frac{1}{d(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^\top (X_i - \bar{X})$. We now define the modified sample kurtosis by

$$\tilde{k}(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \{(X_i - \bar{X})^\top \tilde{S}^{-1} (X_i - \bar{X})\}^2,$$

which is re-written as

$$\tilde{k}(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n} \|X_i - \bar{X}\|^4 / n}{\left(\sum_{i=1}^{n} \|X_i - \bar{X}\|^2 / d(n-1)\right)^2}. \tag{3.12}$$

Under the rotational symmetry assumption, $\tilde{k}(X_1, \ldots, X_n)$ is a consistent estimator of the population multivariate kurtosis (3.4) as stated in the following:

Theorem 7. *Suppose $X_1, \ldots, X_n$ be a random sample from a d-dimensional multivariate distribution with the density f. If $Cov(X_1) = \sigma^2 I_d$ for some $\sigma^2 > 0$, then the modified sample kurtosis $\tilde{k}(X_1, \ldots, X_n)$ satisfies that*

$$\tilde{k}(X_1, \ldots, X_n) \to \kappa(f) \text{ in probability as } n \to \infty.$$

Proof. It is straightforward that s^2 is a consistent estimator of σ^2, thus the theorem follows by Mardia (1970). □

3.4.3.2 Proposed test procedure

In Section 3.4.2, we have investigated the population multivariate kurtosis for rotational symmetric distributions representing the cases of modality. Figure 24 shows that the population multivariate kurtosis is monotonically decreasing from Modality 1 (a single mode) to Modality 3 (modal-ridge). In addition, because we treat Modalities 1 and 2 (a single mode and no mode) as the cases of no major variation, these cases need to be separated from Modality 3 to appropriately fit a great sphere in the PNS procedure. Hence, the test of modality for rotationally symmetric distributions can be performed by using the multivariate kurtosis with hypotheses given by

$$H_{0b''} : \kappa(f) = \kappa(f_\theta) \text{ vs } H_{1b''} : \kappa(f) < \kappa(f_\theta), \tag{3.13}$$

where $\kappa(f_\theta)$ is the population multivariate kurtosis for the ball uniform distribution given in (3.8). From Theorem 7, we know that the modified sample kurtosis is a consistent estimator of population kurtosis for the ball uniform distribution. Therefore, we reject the null hypothesis if the modified sample kurtosis is significantly smaller than a critical value, which is determined by the asymptotic sampling distribution of the modified sample kurtosis investigated in next section.

3.4.3.3 Asymptotic sampling distribution of the modified sample kurtosis

Now, we consider the sampling distribution of the modified sample kurtosis (3.12) when the underlying distribution is assumed to be the ball uniform.

Proposition 8. *Let $Y_1, \ldots, Y_n$ be a random sample from the ball uniform distribution with the radius parameter θ. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\tilde{k}(Y_1, \ldots, Y_n) - \frac{d(d+2)^2}{d+4}\right) \to N\left(0, \frac{128d(d+2)^4}{(d+4)^3(d+6)(d+8)}\right) \text{ in distribution.}$$

Proof. Without loss of generality, we assume $\theta = 1$. By the Taylor series expansion, we have

$$\tilde{k}(Y_1, \ldots, Y_n) = \frac{\bar{R} + O_p(n^{-1})}{\left(\bar{T} + O_p(n^{-1})\right)^2} = \frac{\bar{R} + O_p(n^{-1})}{\bar{T}^2 + O_p(n^{-1})} = \frac{\bar{R}}{\bar{T}^2} + O_p(n^{-1}), \tag{3.14}$$

where $\bar{R} = \frac{1}{n}\sum_{i=1}^n R_i$ with $R_i = (Y_i^\top Y_i)^2$ and $\bar{T} = \frac{1}{n}\sum_{i=1}^n T_i$ with $T_i = (Y_i^\top Y_i)/d$ for $i = 1, \ldots, n$. The detailed calculations for this approximation (3.14)is given in Section 3.7.3.1.

Because R_i's and T_i's are functions of Y_i's, $R_1, \ldots, R_n$ are i.i.d. with the mean $\mu_R := E(R_1) = \frac{d}{d+4}$ and the variance $\sigma_R^2 := Var(R_1) = \frac{16d}{(d+4)^2(d+8)}$, and $T_1, \ldots, T_n$ are also i.i.d. with the mean $\mu_T := E(T_1) = \frac{1}{d+2}$ and the variance $\sigma_T^2 := Var(T_1) = \frac{4}{d(d+2)^2(d+4)}$. Further, if we let $\sigma_{R,T}$ denote the covariance of R_i and T_i, then it is given by $\sigma_{R,T} = \frac{8}{(d+2)(d+4)(d+6)}$. See Section 3.7.3.2 for detailed calculations. By the central limit theorem, as $n \to \infty$,

$$\sqrt{n}\left(\begin{bmatrix} \bar{R} \\ \bar{T} \end{bmatrix} - \begin{bmatrix} \frac{d}{d+4} \\ \frac{1}{d+2} \end{bmatrix}\right) \to N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{16d}{(d+4)^2(d+8)} & \frac{8}{(d+2)(d+4)(d+6)} \\ \frac{8}{(d+2)(d+4)(d+6)} & \frac{4}{d(d+2)^2(d+4)} \end{bmatrix}\right)$$

in distribution. Finally, for the given function $g(\bar{R}, \bar{T}) = \frac{\bar{R}}{\bar{T}^2}$, the multivariate delta method gives us that, as $n \to \infty$,

$$\sqrt{n}\left[g(\bar{R}, \bar{T}) - g(\mu_R, \mu_T)\right] \to N\left(0, \tau^2\right) \text{ in distribution,}$$

where

$$\begin{aligned}\tau^2 &= \left[\frac{\partial g(\mu_R,\mu_T)}{\partial \mu_R}\ \frac{\partial g(\mu_R,\mu_T)}{\partial \mu_T}\right]\begin{bmatrix}\sigma_R^2 & \sigma_{R,T}\\ \sigma_{R,T} & \sigma_T^2\end{bmatrix}\begin{bmatrix}\frac{\partial g(\mu_R,\mu_T)}{\partial \mu_R}\\ \frac{\partial g(\mu_R,\mu_T)}{\partial \mu_T}\end{bmatrix}\\ &= \frac{128d(d+2)^4}{(d+4)^3(d+6)(d+8)}.\end{aligned}$$

□

Proposition 8 says that the modified sample kurtosis (3.12) is asymptotically unbiased and normally distributed. Hence, for the test of modality defined in Section 3.4.3.2, the p-value is obtained accordingly.

3.5 NUMERICAL STUDIES

In this section, we demonstrate the performances of our proposed test of modality for rotationally symmetric distributions by applying the test to the PNS procedure. Our proposed test is applied to the sequential test algorithm of PNS as follows:

Step 1 : The likelihood ratio test for geodesic radius with hypotheses H_{0a} vs H_{1a} in (3.1) is conducted. If H_{0a} is accepted, then the procedure fits a great sphere with $r = \pi/2$ for the current layer and proceed to the next.

Step 2 : If H_{0a} is rejected, then the procedure runs the test of rotational symmetry with hypotheses $H_{0b'}$ vs $H_{1b'}$ in (3.3). If $H_{0b'}$ is rejected, then the procedure uses the fitted small sphere for the current layer and proceed to the next layer.

Step 3 : If $H_{0b'}$ is accepted, then the procedure runs the test of modality with hypotheses $H_{0b''}$ vs $H_{1b''}$ given in (3.13). This test is conducted by using the images of inverse exponential map at the geodesic mean of directional vectors, and by assuming their underlying distribution is the ball uniform. If $H_{0b''}$ is rejected, again the procedure uses the fitted small sphere for the current layer and proceed to the next layer.

Step 4 : If $H_{0b''}$ is accepted, then the procedure fits a great sphere with $r = \pi/2$ for all further subsphere fittings.

Note that the bootstrap von Mises-Fisher test for Case II in Section 3.2.1 is substituted by the test of rotational symmetry (Section 3.3) and our proposed test of modality (Section 3.4.3.2).

We consider several data situations listed below in the simulation study and compare the performances of the proposed approach with other existing methods discussed in Section 3.2. To clearly show the data situations, we assume $d = 2$ so that the data points are distributed on the unit sphere $\mathbb{S}^2$. Figure 25 shows random data examples from below models.

Model 1 : The von Mises-Fisher distribution (vMF; Mardia and Jupp, 2000, p.168)

Model 2 : The tangent space normal distribution with $\Sigma = \sigma^2 I_2$

Model 3 : The tangent space (2-dimensional) ball uniform distribution with the density in (3.6)

Model 4 : The geodesic ball uniform distribution with the density in (3.15)

Model 5 : The small sphere distribution of the second kind (S2; (2.6) in Chapter 2)

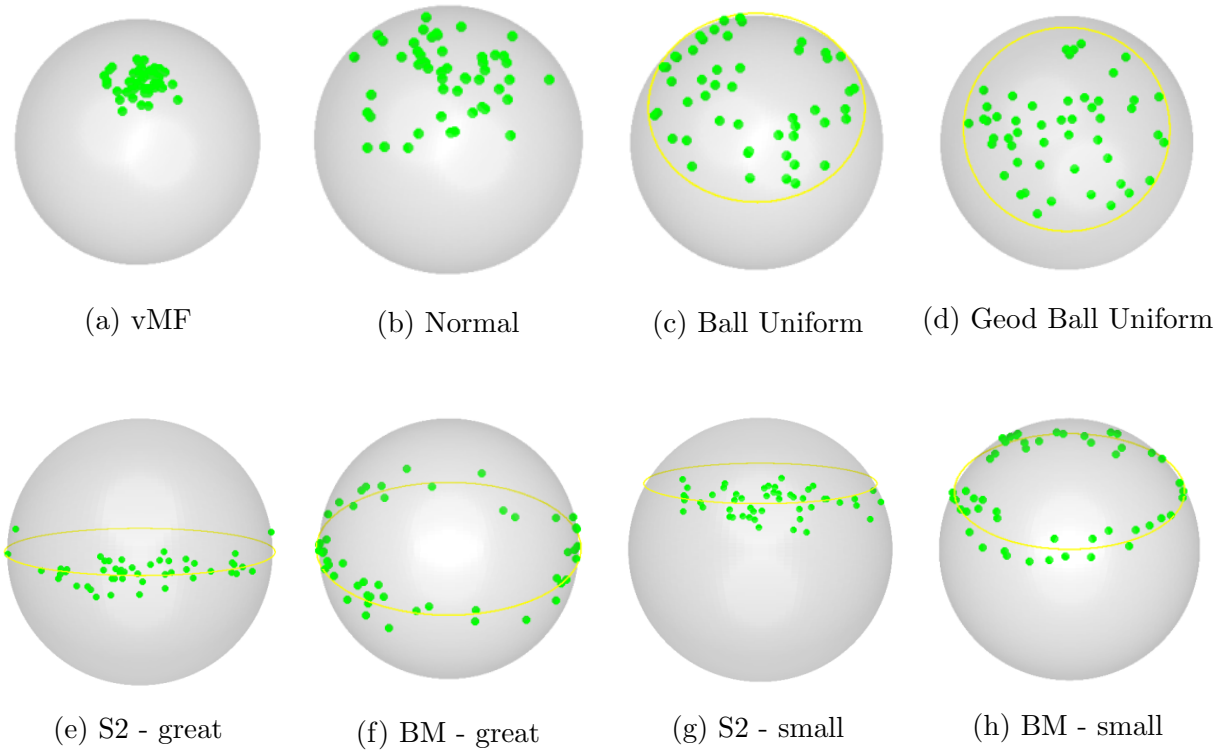
Model 6 : The Bingham-Mardia distribution (BM; Bingham and Mardia, 1978))

The first four models represent the situations that we need to fit a great circle in the PNS procedure because there is no major direction of variation. For the S2 and BM distributions in Models 5 and 6, we consider both small and great circle situations.

The performances of competing methods in the PNS procedure are summarized in the following tables. In these tables, 'PNS' indicates the sequential tests (Section 3.2.1), 'PNS BIC' for the Bayesian information criterion approach (Section 3.2.2), 'Folded Normal' for the folded normal test (Section 3.2.3), 'Tangent Normal' for the test of rotational symmetry by tangent space likelihood ratio test only (Section 3.3). Our proposal is 'TN + Kurtosis', representing Steps 1–4 above, utilizing both the test of rotational symmetry and the test of modality. 'Oracle' indicates the truth for each data situation.

Table 7 shows the result for the sample size $n = 50$ and Table 8 is for larger sample size $n = 200$. We note all methods perform similarly for both $n = 50$ and $n = 200$ cases.

As a measure of performance, the proportion of small sphere fittings among 1000 repetitions is calculated. Because the cases (a) – (d) represent distributions which have no major non-geodesic variation, we prefer a great sphere fit, so lower proportion indicates better performance. For the cases (e) and (f), a random sample is generated from a distribution with a



(a) vMF (b) Normal (c) Ball Uniform (d) Geod Ball Uniform

(e) S2 - great (f) BM - great (g) S2 - small (h) BM - small

Figure 25: Random data examples ($n = 50$) from representing distributions.

geodesic major variation (great sphere). So, in these cases again, lower proportion is better. For the remaining cases (g) and (h), because there is a non-geodesic major variation (small sphere), higher proportion indicates better performance.

Table 7 and 8 confirm that the combination of the test of rotational symmetry and the test of modality shows the best overall performance. As argued by Eltzner et al. (2017), the folded normal test shows better performance than the bootstrap von Mises-Fisher test even when the null distribution is von Mises-Fisher given in the case (a). However, we notice that both tests become poor when their distributional assumptions are not satisfied as in the cases (c) and (d). As expected, the BIC is only effective in distinguishing a geodesic major variation corresponding to cases (e) and (f) from a non-geodesic variation, cases (g) and (h). The weakness of the test of rotational symmetry only is given in the case (h).

As we mentioned before, only with the test of rotational symmetry, a distribution with modal-ridge cannot be separated from other isotropic distributions. Overall, the proposed approach, given in Steps 1–4 above, treats all Models most appropriately, so it seems to alternate existing ones.

3.6 APPENDIX: THE GEODESIC BALL UNIFORM ON $\mathbb{S}^2$

We considered the ball uniform distribution in the tangent space in the development of the testing procedure. However, when the ball uniform distribution is mapped to the sphere, it is no longer a uniform distribution, even though its support is the geodesic ball of radius θ (provided that $\theta < \pi$). For applications to the directional data, it is ideal to consider the geodesic ball uniform distribution, rather than the log-mapped ball uniform distribution. In this section, we discuss the geodesic ball distribution and its multivariate kurtosis when mapped to the tangent space, and argue that the difference between these two distributions are small.

A simple situation that directions on the unit sphere (S^2) show a rotationally symmetric distribution is that those directions are uniformly distributed on a geodesic ball on the sphere, as illustrated in Figure 25 (d). The geodesic ball at μ with the geodesic radius τ is $B_\tau(\mu) = \{x \in \mathbb{S}^2 : \arccos(x^\top \mu) \leq \tau\}$. To define the density of such a local uniform distribution on S^2, we assume the center μ to be the north pole without loss of generality. Then, a random direction X on S^2 can be written by two angles (θ, ϕ), for latitude and longitude, and the density is given by

$$f(\theta, \phi) = \frac{1}{2\pi(1 - \cos\tau)} \sin\theta\,, \quad \theta \in [0, \tau]\,,\ \phi \in [0, 2\pi)\,, \tag{3.15}$$

where $\tau \in (0, \pi/2)$ is the radius of the geodesic ball. Because a simple uniform distribution which is defined on a specifically small region on the unit sphere is of interest, we assumed the radius τ to be less than $\pi/2$, i.e., the region is smaller than a hemisphere.

	Data types							
	Isotropic				Great		Small	
Method	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
PNS	0.050	0.059	0.818	0.822	0.059	0.066	0.963	1.000
PNS BIC	0.999	0.998	1.000	1.000	0.057	0.064	0.963	1.000
Folded Normal	0.002	0.003	0.225	0.166	1.000	1.000	1.000	1.000
Tangent Normal	0.026	0.016	0.003	0.002	0.059	0.002	0.952	0.004
TN + Kurtosis	0.026	0.016	0.107	0.063	0.059	0.056	0.952	0.977
Oracle	0	0	0	0	0	0	1	1

Table 7: Proportions of small sphere fittings with 1000 repetitions; Random samples of size $n = 50$ from (a) vMF (b) Tangent normal (c) Ball uniform (d) Geodesic ball uniform (e) S2 - great (g) BM - great (h) S2 - small (j) BM - small

	Data types							
	Isotropic				Great		Small	
Method	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
PNS	0.056	0.121	1.000	1.000	0.056	0.062	1.000	1.000
PNS BIC	1.000	1.000	1.000	1.000	0.022	0.031	1.000	1.000
Folded Normal	0	0	0.338	0.219	100	100	1.000	1.000
Tangent Normal	0.017	0.019	0	0.002	0.056	0	1.000	0
TN + Kurtosis	0.017	0.019	0.052	0.026	0.056	0.062	1.000	1.000
Oracle	0	0	0	0	0	0	1	1

Table 8: Proportions of small sphere fittings with 1000 repetitions; Random samples of size $n = 200$ from (a) vMF (b) Tangent normal (c) Ball uniform (d) Geodesic ball uniform (e) S2 - great (g) BM - great (h) S2 - small (j) BM - small

Let $Y \in \mathbb{R}^2$ be the image of inverse exponential map of the direction X onto the tangent space at the center μ. Then, the density of Y is obtained by

$$f_\tau(y) = \frac{1}{2\pi(1-\cos\tau)} \frac{\sin\|y\|}{\|y\|}, \tag{3.16}$$

where $\|y\| = \sqrt{y_1^2 + y_2^2} \in [0, \tau]$ for $y = (y_1, y_2)^\top$.

The measure of multivariate kurtosis for the 'tangent space-mapped' X, or Y, is

$$\kappa(f_\tau) = 4(1-\cos\tau)\frac{c_1(\tau)}{c_2(\tau)}, \tag{3.17}$$

where

$$\begin{aligned} c_1(\tau) &= -\tau^4\cos\tau + 4\tau^3\sin\tau + 12\tau^2\cos\tau - 24\cos\tau - 24\tau\sin\tau + 24, \\ c_2(\tau) &= (-\tau^2\cos\tau + 2\tau\sin\tau + 2\cos\tau - 2)^2. \end{aligned}$$

It is challenging to obtain the generalized function of kurtosis for a higher dimension, $\mathbb{S}^d$, $d \geq 3$.

The measure of multivariate kurtosis for the geodesic ball uniform distribution on $\mathbb{S}^2$ depends on the radius τ. In contrast, the kurtosis for the ball uniform distribution does not depend on the radius of its support. However, when τ is small enough, both the kurtosis and the density of geodesic ball uniform are similar to those of ball uniform. As τ becomes smaller ($\tau \to 0$), $\sin\|\mathbf{y}\| \approx \|\mathbf{y}\|$, which leads the density (3.16) to the ball uniform density (3.6). Hence, we consider an approximation by using the ball uniform distribution.

3.7 APPENDIX: TECHNICAL DETAILS

3.7.1 Proof of Lemma 5

All the expectations stated in Lemma 5 can be calculated by considering the spherical coordinate transformation, $U = (U_1, \ldots, U_d)^\top = (r\cos\phi_1, \ldots, r\prod_{i=1}^{d-1}\sin\phi_i)^\top$ and $du_1 \cdots du_d = r^{d-1}\sin^{d-2}\phi_1 \cdots \sin\phi_{d-2} dr d\phi_1 \cdots d\phi_{d-1}$ with $r \in [\theta_1, \theta_2]$, $\phi_i \in [0, \pi)$ for $i = 1, \ldots, d-2$, and $\phi_{d-1} \in [0, 2\pi)$. With this transformation, we first obtain the following equation,

$$\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\sin^{d-2}\phi_1 \cdots \sin\phi_{d-2} d\phi_1 \cdots d\phi_{d-1} = \frac{(V_{\theta_2} - V_{\theta_1})d}{\theta_2^d - \theta_1^d}, \tag{3.18}$$

which is obtained by the fact that $\int f_{\theta_1,\theta_2} du = 1$.

From any integration table, we can easily find a trigonometric integration formula

$$\int \sin^n u \cos^m u du = \frac{\sin^{n+1} u \cos^{m-1} u}{n+m} + \frac{m-1}{n+m}\int \sin^n u \cos^{m-2} u du,$$

for nonnegative integer n and m, which is useful for the calculation of expectations. As contained in (3.18), in our calculations, n is always greater than or equal to 1, which makes the first term in the right hand side of the formula 0. Recall $\sin^n(0) = \sin^n(\pi) = \sin^n(2\pi) = 0$. So, when $m \geq 1$, this formula returns two different results, which are

$$\int \sin^n u \cos^m u du = \begin{cases} \frac{(m-1)(m-3)\cdots 2}{(n+m)(n+m-2)\cdots(n+3)}\int \sin^n u \cos u du & , m \text{ is odd} \\ \frac{(m-1)(m-3)\cdots 1}{(n+m)(n+m-2)\cdots(n+2)}\int \sin^n u du & , m \text{ is even.} \end{cases}$$

In fact, it finally returns 0 when m is odd because $\int_0^\pi \sin^n u \cos u du = \sin^{n+1} u/(n+1)|_0^\pi = 0$. Hence, we have

$$E(U_j^s) = 0$$

when s is odd. When s is even, we can show that, for $j = 1$,

$$\begin{aligned}
E(U_1^s) &= \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\int_{\theta_1}^{\theta_2}\frac{1}{V_{\theta_2}-V_{\theta_1}}(r^s\cos^s\phi_1)r^{d-1}\sin^{d-2}\phi_1\cdots\sin\phi_{d-2}drd\phi_1\cdots d\phi_{d-1}\\
&= \frac{\theta_2^{d+s}-\theta_1^{d+s}}{V_{\theta_2}-V_{\theta_1}}\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}(\cos^s\phi_1\sin^{d-2}\phi_1)\cdots\sin\phi_{d-2}d\phi_1\cdots d\phi_{d-1}\\
&= \frac{\theta_2^{d+s}-\theta_1^{d+s}}{V_{\theta_2}-V_{\theta_1}}\frac{(s-1)(s-3)\cdots 1}{(d+s)(d+s-2)\cdots d}\int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\sin^{d-2}\phi_1\cdots\sin\phi_{d-2}d\phi_1\cdots d\phi_{d-1}\\
&= \frac{\theta_2^{d+s}-\theta_1^{d+s}}{V_{\theta_2}-V_{\theta_1}}\frac{(s-1)(s-3)\cdots 1}{(d+s)(d+s-2)\cdots d}\frac{(V_{\theta_2}-V_{\theta_1})d}{\theta_2^d-\theta_1^d}\\
&= d\frac{(\theta_2^{d+s}-\theta_1^{d+s})}{(\theta_2^d-\theta_1^d)}\frac{(s-1)!!}{(d+s)(d+s-2)\cdots d},
\end{aligned}$$

where $c!!$ is the double factorial defined by $c(c-2)\cdots 1$. Because U_j's have the same marginal distributions for all $j = 1,\ldots,d$, we can easily show that $E(U_j^s)$'s are the same for all $j = 1,\ldots,d$. (We skip the details here.)

The second and third parts of Lemma 5 are also proven by similar way. For the second part, the expectation is calculated as follows:

$$\begin{aligned}
E(U_j\prod_{j'\neq j}U_{j'}^{s_{j'}}) &= E(U_1\prod_{j'=2}^{d}U_{j'}^{s_{j'}})\\
&= \int\cdots\int\frac{1}{V_{\theta_2}-V_{\theta_1}}(r\cos\phi_1)(\prod_{j'\neq j}U_{j'}^{s_{j'}})r^{d-1}\sin^{d-2}\phi_1\cdots\sin\phi_{d-2}drd\phi_1\cdots d\phi_{d-1}\\
&= \cdots\left(\int_0^{\pi}\cos\phi_1\sin^{d-2+\sum_{j'=2}^{d}s_{j'}}d\phi_1\right)\cdots\\
&= 0.
\end{aligned}$$

Again, we omit the details but we can easily show that the results are the same for all $j = 1,\ldots,d$.

For the third part, it is calculated by

$$\begin{aligned}
E(U_1^2U_2^2) &= \int\cdots\int \frac{1}{V_{\theta_2}-V_{\theta_1}}(r^2\cos^2\phi_1)(r^2\sin^2\phi_1\cos^2\phi_2)r^{d-1}\sin^{d-2}\phi_1\cdots \\
&= \frac{\theta_2^{d+4}-\theta_1^{d+4}}{V_{\theta_2}-V_{\theta_1}}\int\cdots\int(\cos^2\phi_1\sin^d\phi_1)(\cos^2\phi_2\sin^{d-3}\phi_2)\cdots \\
&= \frac{\theta_2^{d+4}-\theta_1^{d+4}}{V_{\theta_2}-V_{\theta_1}}\left(\frac{d-1}{(d+2)d}\int_0^\pi \sin^{d-2}\phi_1 d\phi_1\right)\left(\frac{1}{d-1}\int_0^\pi \sin^{d-3} d\phi_2\right)\cdots \\
&= \frac{\theta_2^{d+4}-\theta_1^{d+4}}{V_{\theta_2}-V_{\theta_1}}\frac{1}{d(d+2)}\frac{(V_{\theta_2}-V_{\theta_1})d}{\theta_2^d-\theta_1^d} \\
&= \frac{(\theta_2^{d+4}-\theta_1^{d+4})}{(\theta_2^d-\theta_1^d)}\frac{1}{(d+4)(d+2)}.
\end{aligned}$$

3.7.2 Proof of Propositoin 6

First, for a convolution form Z with a given constant $c \in [0,1]$, we know that $E(Z) = 0$ because, for both X and Y, we assumed rotationally symmetric distributions centered at the origin. The population variance-covariance matrix is given by

$$\begin{aligned}
\Sigma_Z &= E\left[(Z-E(Z))(Z-E(Z))^\top\right] \\
&= E\left[((1-c)X+cY)((1-c)X+cY)^\top\right] \\
&= E\left[(1-c)^2XX^\top + c(1-c)XY^\top + c(1-c)YX^\top + c^2YY^\top\right] \\
&= (1-c)^2E(XX^\top) + c^2E(YY^\top) \\
&= (1-c)^2\sigma^2I_d + c^2\frac{\theta^2}{d+2}I_d.
\end{aligned}$$

The required moments for X are well-known and the ones for Y are obtained by Lemma 5. Then, the population multivariate kurtosis for Z is calculated by

$$\begin{aligned}
\kappa(f_{\tau,c}) &= E\left[((Z-E(Z))^\top\Sigma_Z^{-1}(Z-E(Z)))^2\right] \\
&= \left((1-c)^2\sigma^2I_d + c^2\frac{\theta^2}{d+2}I_d\right)^{-2}E\left[(Z^\top Z)^2\right] \\
&= \left((1-c)^2\sigma^2I_d + c^2\frac{\theta^2}{d+2}I_d\right)^{-2}\left[dE(Z_j^4) + d(d-1)E(Z_j^2Z_{j'}^2)\right].
\end{aligned}$$

By the independence of X and Y, $E(Z_j^4)$ is calculated by

$$\begin{aligned} E(Z_j^4) &= E\left[((1-c)X_j + cY_j)^4\right] \\ &= E\left[(1-c)^4 X_j^4 + c^4 Y_j^4 + 4(1-c)^3 c X_j^3 Y_j + 4(1-c)c^3 X_j Y_j^3 + 6(1-c)^2 c^2 X_j^2 Y_j^2\right] \\ &= (1-c)^4 E(X_j^4) + c^4 E(Y_j^4) + 6(1-c)^2 c^2 E(X_j^2) E(Y_j^2) \\ &= (1-c)^4 \frac{3\theta^4}{(d+2)(d+4)} + c^4 3\sigma^4 + 6(1-c)^2 c^2 \frac{\theta^2 \sigma^2}{d+2}, \end{aligned}$$

and $E(Z_j^2 Z_{j'}^2)$ is calculated by

$$\begin{aligned} E(Z_j^2 Z_{j'}^2) &= E\left[((1-c)X_j + cY_j)^2((1-c)X_{j'} + cY_{j'})^2\right] \\ &= E\left[((1-c)^2 X_j^2 + 2(1-c)cX_j Y_j + c^2 Y_j)((1-c)^2 X_{j'}^2 + 2(1-c)cX_{j'} Y_{j'} + c^2 Y_{j'})\right] \\ &= (1-c)^4 E(X_j^2 X_{j'}^2) + c^2(1-c)^2 E(Y_j^2) E(X_{j'}^2) + (1-c)^2 c^2 E(X_j^2) E(Y_{j'}^2) + c^4 E(Y_j^2 Y_{j'}^2) \\ &= (1-c)^4 \frac{\theta}{(d+2)(d+4)} + 2(1-c)^2 c^2 \frac{\theta^2 \sigma^2}{d+2} + (1-c)^4 \sigma^4. \end{aligned}$$

Hence, the population multivariate kurtosis for Z is obtained by

$$\kappa(f_{\tau,c}) = \left((1-c)^2 \sigma^2 I_d + c^2 \frac{\theta^2}{d+2} I_d\right)^{-2} \left[(1-c)^4 \frac{d\theta^4}{d+4} + c^4 d(d+2)\sigma^4 + 2(1-c)^2 c^2 d\theta^2 \sigma^2\right],$$

and, by letting $\tau = \theta/\sigma$, (3.10) follows.

Second, for a mixture Z with a given $p = P(c=1) \in [0,1]$ for $c \in \{0,1\}$, the required moments are calculated by follows: for $j = 1, \ldots, d$ and $j \neq j'$,

$$\begin{aligned} &E(Z_j) = E(Z_j|c=1) + E(Z_j|c=0) = (1-p)E(X_j) + pE(Y_j) = 0, \\ &E(Z_j^2) = (1-p)E(X_j^2) + pE(Y_j^2) = (1-p)\sigma^2 + p\frac{\theta^2}{d+2}, \\ &E(Z_j Z_{j'}) = (1-p)E(X_j X_{j'}) + pE(Y_j Y_{j'}) = 0, \\ &E(Z_j^2 Z_{j'}^2) = (1-p)E(X_j^2 X_{j'}^2) + pE(Y_j^2 Y_{j'}^2) = (1-p)\sigma^4 + p\frac{\theta^4}{(d+2)(d+4)}, \\ &E(Z_j^4) = (1-p)E(X_j^4) + pE(Y_j^4) = (1-p)3\sigma^4 + p\frac{3\theta^4}{(d+2)(d+4)}. \end{aligned}$$

From these moments, we obtain the variance-covariance matrix,

$$\Sigma_Z = \left((1-p)\sigma^2 + p\frac{\theta^2}{d+2}\right) I_d,$$

and then, the population multivariate kurtosis for a mixture Z is given by

$$\begin{aligned}\kappa(f_{\tau,p}) &= E\left[((Z-E(Z))^\top \Sigma_Z^{-1}(Z-E(Z)))^2\right] \\ &= \left((1-p)\sigma^2 + p\frac{\theta^2}{d+2}\right)^{-2} E\left[(Z^\top Z)^2\right] \\ &= \left((1-p)\sigma^2 + p\frac{\theta^2}{d+2}\right)^{-2} \left[dE(Z_j^4) + d(d-1)E(Z_j^2 Z_{j'}^2)\right],\end{aligned}$$

which, by letting $\tau = \theta/\sigma$, turns out to be (3.11).

3.7.3 Details in proof of Proposition 8

3.7.3.1 Approximation of the modified sample kurtosis in (3.14)

Recall that the modified sample kurtosis (3.12) is the ratio of two multivariate central moments, $\frac{1}{n}\sum_{i=1}^n \|Y_i - \bar{Y}\|^4$ and $\left(\frac{1}{d(n-1)}\sum_{i=1}^n \|Y_i - \bar{Y}\|^2\right)^2$. First, if the numerator is denoted by a function of $\bar{Y}$, $f(\bar{Y})$, then the Taylor expansion of $f(\bar{Y})$ around $\bar{Y} = 0$ is

$$f(\bar{Y}) = f(0) + (\bar{Y}-0)^\top \nabla f(0) + \frac{1}{2}(\bar{Y}-0)^\top \nabla^2 f(0)(\bar{Y}-0) + \text{higher order terms}.$$

Here, for $\nabla f(0)$, we can show that

$$\begin{aligned}\|\nabla f(0)\| &= \|(-4)\frac{1}{n}\sum_{i=1}^n (Y_i^\top Y_i)Y_i\| \\ &\le \frac{4}{n}\sum_{i=1}^n \|(Y_i^\top Y_i)Y_i\| \\ &= \frac{4}{n}\sum_{i=1}^n (Y_i^\top Y_i)\|Y_i\| \\ &= \frac{4}{n}\sum_{i=1}^n (Y_i^\top Y_i)^{3/2},\end{aligned}$$

where the inequality is obtained by the triangular inequality. By letting $W_i = 4(Y_i^\top Y_i)^{3/2}$, the last line becomes the sample mean of W_i's. Because W_i's are bounded, we have $\|\nabla f(0)\| = O_p(n^{-1/2})$. Hence, $\bar{R} + O_p(n^{-1})$ is obtained in (3.14) with $\bar{R} = f(0) = \frac{1}{n}\sum_{i=1}^n (Y_i^\top Y_i)^2$.

Second, let us denote the denominator by $\{g(\bar{Y})\}^2$, where $g(\bar{Y}) = \frac{1}{d(n-1)} \sum_{i=1}^{n} \|Y_i - \bar{Y}\|^2$, the pooled sample variance. Then, by similar way, we can show that

$$g(\bar{Y}) = g(0) + O_p(n^{-1}).$$

Furthermore, because $g(0)$ satisfies that

$$\begin{aligned} g(0) &= \frac{1}{dn} \sum_{i=1}^{n} Y_i^\top Y_i - \frac{1}{dn(n-1)} \sum_{i=1}^{n} Y_i^\top Y_i \\ &= \frac{1}{dn} \sum_{i=1}^{n} Y_i^\top Y_i + O_p(n^{-1}), \end{aligned}$$

$\bar{T} + O_p(n^{-1})$ in (3.14) is obtained with $\bar{T} = \frac{1}{dn} \sum_{i=1}^{n} Y_i^\top Y_i$.

3.7.3.2 Calculation of the expected values and variances of R_i and T_i

As stated before, we assume $\theta = 1$, without loss of generality. First, the expected value of R_i is calculated by

$$\begin{aligned} E(R_i) = E\left[(Y_i^\top Y_i)^2\right] &= E\left[(Y_{1i}^2 + \cdots + Y_{di}^2)^2\right] \\ &= E\left[\sum_{j=1}^{d} Y_{ji}^4 + \sum\sum_{j \neq j'} Y_{ji}^2 Y_{j'i}^2\right] \\ &= dE(Y_{1i}^4) + d(d-1)E(Y_{1i}^2 Y_{2i}^2) \\ &= d\frac{3}{(d+2)(d+4)} + d(d-1)\frac{1}{(d+2)(d+4)} \\ &= \frac{d}{d+4}, \end{aligned}$$

where $E(Y_{1i}^4)$ and $E(Y_{1i}^2 Y_{2i}^2)$ are obtained by Lemma 5. Then, the variance of R_i is obtained by

$$\begin{aligned} Var(R_i) = Var\left[(Y_i^\top Y_i)^2\right] &= E\left[(Y_i^\top Y_i)^4\right] - \left\{E\left[(Y_i^\top Y_i)^2\right]\right\}^2 \\ &= \frac{d}{d+8} - \frac{d^2}{(d+4)^2} \\ &= \frac{16d}{(d+4)^2(d+8)}, \end{aligned}$$

where $E\left[(Y_i^\top Y_i)^4\right]$ is calculated by

$$\begin{aligned}
E\left[(Y_i^\top Y_i)^4\right] &= E\left[(Y_{1i}^2+\cdots+Y_{di}^2)^4\right] \\
&= E\left[\sum_{j=1}^{d} Y_{ji}^8 + \sum_{j\neq j'} Y_{ji}^6 Y_{j'i}^2 + \sum_{j\neq k\neq l} Y_{ji}^4 Y_{ki}^2 Y_{li}^2 + \sum_{j\neq k\neq l\neq m} Y_{ji}^2 Y_{ki}^2 Y_{li}^2 Y_{mi}^2\right] \\
&= dE(Y_{1i}^8) + 4d(d-1)E(Y_{1i}^6 Y_{2i}^2) + 3d(d-1)E(Y_{1i}^4 Y_{2i}^4) \\
&\quad +6d(d-1)(d-2)E(Y_{1i}^4 Y_{2i}^2 Y_{3i}^2) + d(d-1)(d-2)(d-3)E(Y_{1i}^2 Y_{2i}^2 Y_{3i}^2 Y_{4i}^2).
\end{aligned}$$

These expectations on the last line can be calculated by following Lemma 5 and its proof, which returns

$$\begin{aligned}
E(Y_{1i}^8) &= \frac{105}{(d+2)(d+4)(d+6)(d+8)}, \\
E(Y_{1i}^6 Y_{2i}^2) &= \frac{15}{(d+2)(d+4)(d+6)(d+8)}, \\
E(Y_{1i}^4 Y_{2i}^4) &= \frac{9}{(d+2)(d+4)(d+6)(d+8)}, \\
E(Y_{1i}^4 Y_{2i}^2 Y_{3i}^2) &= \frac{3}{(d+2)(d+4)(d+6)(d+8)}, \\
E(Y_{1i}^2 Y_{2i}^2 Y_{3i}^2 Y_{4i}^2) &= \frac{1}{(d+2)(d+4)(d+6)(d+8)}.
\end{aligned}$$

Hence, by plugging these in above, we get $E\left[(Y_i^\top Y_i)^4\right] = \frac{d}{d+8}$.

The expected value and variance of T_i are obtained by using Lemma 5 and above calculations directly,

$$E(T_i) = \frac{1}{d}E(Y_i^\top Y_i) = E(Y_{1i}^2) = \frac{1}{d+2},$$

and

$$\begin{aligned}
Var(T_i) &= \frac{1}{d^2}\left\{E\left[(Y_i^\top Y_i)^2\right] - \left[E(Y_i^\top Y_i)\right]^2\right\} \\
&= \frac{1}{d^2}\left(\frac{d}{d+4} - \frac{d^2}{(d+2)^2}\right) \\
&= \frac{4}{d(d+2)^2(d+4)}.
\end{aligned}$$

BIBLIOGRAPHY

Banerjee, A., Dhilon, I. S., Ghosh, J., and Sra, S. (2005), "Clustering on the unit hypersphere using von Mises-Fisher distributions," *Journal of Machine Learning Research*, 6, 1345–1382.

Bingham, C. and Mardia, K. V. (1978), "A small circle distribution on the sphere," *Biometrika*, 65, 379–389.

Browne, M. W. (1967), "On oblique Procrustes rotation," *Psychometrika*, 32, 125–132.

Cootes, T. F., Taylor, C., Cooper, D., and Graham, J. (1992), "Training Models of Shape From Sets of Examples," in *Proceedings of British Machine Vision Conference*, eds. Hogg, D. and Boyle, R., Berlin: Springer-Verlag, pp. 9–18.

Dryden, I. and Mardia, K. V. (1998), *Statistical Shape Analysis*, Chichester: Wiley.

Eltzner, B., Huckemann, S., and Mardia, K. V. (2017), "Torus Principal Component Analysis with an Application to RNA Structures," *Annals of Applied Statistics*, ISSN 1932-6157 (In Press).

Eltzner, B., Jung, S., and Huckemann, S. (2015), "Dimension reduction on polyspheres with application to skeletal representations," *Geometric Science of Information 2015 proceedings*, 22–29.

Fletcher, P. T., Lu, C., M., P. S., and Joshi, S. (2004), "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Tans. Med. Imag.*, 23, 995–1005.

Gray, N. H., Geiser, P. A., and Geiser, J. R. (1980), "On the least-squares fit of small and great circles to spherically projected orientation data," *Journal of the International Association for Mathematical Geology*, 12, 173–184.

Hoff, P. D. (2009), "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data," *Journal of Computational and Graphical Statistics*, 18, 438–456.

Huckemann, S., Hotz, T., and Munk, A. (2010), "Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions," *Statistica Sinica*, 20, 1–58.

Huckemann, S. and Ziezold, H. (2006), "Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces," *Adv. Appl. Prob.*, 38, 299–319.

Joshi, S., Pizer, S., Fletcher, P. T., Yushkevich, P., Thall, A., and Marron, J. S. (2002), "Multiscale deformable model segmentation and statistical shape analysis using medial descriptions," *IEEE transactions on medical imaging*, 21, 538–50.

Jung, S., Dryden, I. L., and Marron, J. S. (2012), "Analysis of Principal Nested Spheres," *Biometrika*, 99, 551–568.

Jung, S., Foskey, M., and Marron, J. S. (2011), "Principal arc analysis on direct product manifolds," *The Annals of Applied Statistics*, 5, 578–603.

Jupp, P. (2005), "Sobolev tests of goodness of fit of distributions on compact Riemannian manifolds," *The Annals of Statistics*, 2957–2966.

— (2008), "Data-driven Sobolev tests of uniformity on compact Riemannian manifolds," *The Annals of Statistics*, 1246–1260.

Kenobi, K., Dryden, I. L., and Le, H. (2010), "Shape curves and geodesic modelling," *Biometrika*, 97, 567–584.

Kent, J. T. (1982), "The Fisher-Bingham distribution on the sphere," *Journal of the Royal Statistical Society. Series B (Methodological)*, 71–80.

Kume, A. and Sei, T. (2018), "On the exact maximum likelihood inference of Fisher–Bingham distributions using an adjusted holonomic gradient method," *Statistics and Computing*, 28, 835–847.

Kume, A. and Wood, A. T. A. (2005), "Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants," *Biometrika*, 92, 465–476.

Kurtek, S., Ding, Z., Klassen, E., and Srivastava, A. (2011), "Parameterization-Invariant Shape Statistics and Probabilistic Classification of Anatomical Surfaces," in *Information Processing in Medical Imaging*, Berlin: Springer, vol. 6801, pp. 147–158.

Lee, M. H. (2007), "Continuum Direction Vectors in High Dimensional Low Sample Size Data," Ph.D. thesis.

Mardia, K. (1970), "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, 57, 519–530.

— (1975), "Statistics of directional data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 37, 349–393.

Mardia, K. V. and Gadsden, R. J. (1977), "A Small Circle of Best Fit for Spherical Data and Areas of Vulcanism," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26, 238–245.

Mardia, K. V., Hughes, G., Taylor, C. C., and Singh, H. (2008), "A multivariate von Mises distribution with applications to bioinformatics," *Canadian Journal of Statistics*, 36, 99–109.

Mardia, K. V. and Jupp, P. E. (2000), *Directional Statistics*, vol. 28 of *Wiley series in probability and statistics*, Wiley.

Pierrynowski, M., Costigan, P., Maly, M., and Kim, P. (2010), "Patients With Osteoarthritic Knees Have Shorter Orientation and Tangent Indicatrices During Gait," *Clinical Biomechanics*, 25, 237–141.

Pizer, S. M., Jung, S., Goswami, D., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., and Marron, J. S. (2013), "Nested Sphere Statistics of Skeletal Models," in *Innovations for Shape Analysis: Models and Algorithms*, eds. Breu, M., Bruckstein, A., and Maragos, P., New York: Springer, pp. 93–115.

Rivest, L. P. (1999), "Some Linear Model Techniques for Analyzing Small-Circle Spherical Data," *Canadian Journal of Statistics*, 27, 623–638.

Schulz, J., Jung, S., Huckemann, S., Pierrynowski, M., Marron, J. S., and Pizer, S. M. (2015), "Analysis of Rotational Deformations from Directional Data," *Journal of Computational and Graphical Statistics*, 24, 539–560.

Shing, H., Hnizdo, V., and Demchuk, E. (2002), "Probabilistic Model for Two Dependent Circular Variables," *Biometrika*, 89, 719–723.

Siddiqi, K. and Pizer, S. (2008), *Medial Representation: Mathematics, Algorithms and Applications*, Springer.

Székely, G. J. and Rizzo, M. L. (2013), "Energy statistics: A class of statistics based on distances," *Journal of statistical planning and inference*, 143, 1249–1272.